



Marcelo Cascardo Cardoso

**Determinação em Tempo Real dos Riscos de
Desligamentos em Linhas de Transmissão devido a
Descargas Atmosféricas**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Engenharia Elétrica.

Orientadora: Prof.^a Marley Maria Bernardes Rebuzzi Vellasco

Rio de Janeiro
Outubro de 2017



Marcelo Cascardo Cardoso

**Determinação em Tempo Real dos Riscos de
Desligamentos em Linhas de Transmissão devido a
Descargas Atmosféricas**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof.^a Marley Maria Bernardes Rebuszi Vellasco
Orientadora
Departamento de Engenharia Elétrica – PUC-Rio

Prof.^a Karla Tereza Figueiredo Leite
Co-orientadora
UERJ

Prof. Delberis Araujo Lima
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Milton Brown Do Coutto Filho
UFF

Prof. Paulo Gomes
ONS

Prof. Márcio da Silveira Carvalho
Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Rio de Janeiro, 02 de outubro de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Marcelo Cascardo Cardoso

Graduou-se em Engenharia Elétrica pela Universidade Federal do Rio de Janeiro (UFRJ) em 1992. Pós graduou-se em Engenharia de Sistemas e Computação pela Universidade do Estado do Rio de Janeiro (UERJ) em 2005 e MBA em Administração e Gerência do Setor Elétrico pela PUC-Rio em 2006. Trabalhou como Engenheiro no CEPEL - Centro de Pesquisas da Eletrobrás e em FURNAS Centrais Elétricas desde 1993. Em 1999 foi para o Operador Nacional do Sistema Elétrico (ONS), onde desde 2007 assumiu a Gerência de Sistemas de Supervisão e Controle do Centro Regional de Operação Sudeste, trabalhando com sistemas SCADA (Supervisory Control and Data Acquisition), EMS (Energy Management System) e GIS (Sistemas de Informações Geográficas).

Ficha Catalográfica

Cardoso, Marcelo Cascardo

Determinação em tempo real dos riscos de desligamentos em linhas de transmissão devido a descargas atmosféricas / Marcelo Cascardo Cardoso ; orientadora: Marley Maria Bernardes Rebuzzi Vellasco ; co-orientadora: Karla Tereza Figueiredo Leite. – 2017.

186 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2017.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Perturbações no sistema elétrico. 3. Interrupção de energia. 4. Operação do sistema interligado. 5. Raios. 6. Descargas atmosféricas. I. Vellasco, Marley Maria Bernardes Rebuzzi. II. Leite, Karla Tereza Figueiredo. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Agradecimentos

Às minhas orientadoras Marley Vellasco e Karla Figueiredo pela dedicação, paciência e confiança deposita ao longo desta jornada. Obrigado por me receberem, acreditado no trabalho e me incentivarem permanentemente. Sem vocês esse objetivo não seria atingido.

Aos meus pais, pela motivação e formação que me proporcionaram.

À minha esposa e aos meus filhos pela compreensão e paciência ao longo deste caminho.

Aos vários colegas do Operador Nacional do Sistema Elétrico (ONS), que contribuíram e me estimularam a elaborar este trabalho.

Resumo

Cardoso, Marcelo Cascardo; Vellasco, Marley Maria Bernardes Rebuzzi (Orientadora); Leite, Karla Tereza Figueiredo (Co-orientadora). **Determinação em Tempo Real dos Riscos de Desligamentos em Linhas de Transmissão devido a Descargas Atmosféricas**. Rio de Janeiro, 2017. 186p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

As descargas atmosféricas são de grande importância para o setor elétrico, sendo frequentemente responsáveis por desligamentos de linhas de transmissão, que podem desencadear uma sequência de eventos que levem o sistema elétrico interligado ao colapso. As longas extensões de linhas de transmissão, expostas a intemperes climáticas, determinam uma probabilidade significativa de incidência direta de descargas atmosféricas nestes equipamentos. Devido ao caráter estratégico das linhas para o fornecimento de energia e a constatação de que descargas atmosféricas estão entre as principais causas de desligamentos, torna-se importante o estudo do comportamento das descargas atmosféricas, antes do instante da ocorrência do desligamento das linhas de transmissão, para compreender os padrões característicos potenciais causadores destes desligamentos. Os estudos encontrados atualmente estão orientados na eficiência das redes de detecção de descargas atmosféricas e na identificação de condições climáticas que indiquem a ocorrência de raios de forma preditiva, sem correlação a ocorrências em linhas de transmissão. Assim, essa dissertação consiste na determinação do risco de desligamentos de linhas de transmissão por descargas atmosféricas, visando fornecer informações antecipadas e possibilitar ações operativas para manter a segurança do sistema elétrico. O modelo desenvolvido nesse estudo, denominado Risco de Desligamentos de Linhas de Transmissão por Raios (RDLR), é composto de dois módulos principais, sendo o primeiro o agrupamento do conjunto amostral de descargas atmosféricas, realizado através de um método baseado em densidade. Nesse módulo, os ruídos são eliminados de forma eficiente e são formados grupos representativos de descargas atmosféricas. O segundo módulo consiste em uma etapa classificatória, baseado em redes neurais artificiais para identificar padrões de grupos de descargas que representem riscos de desligamentos de linhas de

transmissão. Visando a otimização do modelo, foi aplicado um método de seleção das variáveis, através de componentes principais, para determinar aquelas que mais contribuem na caracterização desses eventos. O modelo RDLR foi testado com dados reais dos registros de desligamentos de linhas de transmissão, associado a outro banco com dados reais contendo milhões de registros de descargas atmosféricas oriundos das redes de detecção de raios, sendo obtidos excelentes resultados na determinação dos riscos de desligamentos de linhas de transmissão por descargas atmosféricas.

Palavras-chave

Perturbações no Sistema Elétrico; Interrupção de Energia; Operação do Sistema Interligado; Raios; Descargas Atmosféricas; Desligamentos em Linhas de Transmissão; Sistemas Inteligentes; Redes Neurais MLP; Aprendizado de Máquina; Big Data; DBScan; Análise de Componentes Principais (PCA).

Abstract

Cardoso, Marcelo Cascardo; Vellasco, Marley Maria Bernardes Rebuzzi (Advisor); Leite, Karla Tereza Figueiredo (Co-advisor). **Real-time Risks Determination of Transmission Lines Outage by Lightnings**. Rio de Janeiro, 2017. 186p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Atmospheric discharges are of great importance to power systems, and are often responsible for outages of transmission lines, which can trigger a sequence of events that leads to a system collapse. The long extensions of transmission lines, exposed to climatic conditions, create significant probability of direct incidence of atmospheric discharges in these equipments. Due to the strategic nature of power supply lines and the fact that atmospheric discharges are among the main causes of outages, it is important to study atmospheric discharges characteristics before failure of transmission lines and understand patterns that are responsible for interruptions. Current studies focus on efficiency of lightning detection networks and on identification of climatic conditions that indicate lightning occurrence in a predictive approach, without any correlation with transmission lines outages. Therefore, this thesis consists on real-time risk determination of transmission lines outage by lightning, providing early information to enabling operational procedures for power system safety. The proposed model, named Transmission Lines Outage Risk by Lightning (TLORL) is composed of two main modules: Atmospheric Discharge Data Clustering and Classification. In the atmospheric discharges data-clustering module, performed by a density-based method, the outages are efficiently eliminated and representative groups of atmospheric discharges are formed. The second module consists of a classification step, based on artificial neural networks, to identify patterns of discharges groups that represent risks to cause transmission lines outages. Aiming at improving the proposed model, principal components analysis (PCA) was applied to determine the input variables that most contribute to the events characterization. The TLORL model was tested with real data transmission line outages, associated to another database with millions lightning

records from the detection networks, producing excellent results of transmission lines outages caused by atmospheric discharges.

Keywords

Power System Disturbance; Power Interruption; Interconnected System Operation; Lightning; Atmospheric Discharges; Transmission Lines; Intelligent Systems; Artificial Neural Network; MLP; Machine Learning; Big Data; DBScan; Principal Component Analysis (PCA).

Sumário

1 Introdução	16
1.1. Motivação	16
1.2. Objetivos	17
1.3. Descrição do Trabalho	17
1.4. Organização do Trabalho	20
2 Desligamentos de Linhas de Transmissão por Descargas Atmosféricas	22
2.1. O Fenômeno da Descarga Atmosférica	22
2.1.1. Conceitos Básicos	22
2.1.2. Tipos de Descargas Atmosféricas	23
2.1.3. Sistema de Detecção e Localização de Descargas Atmosféricas	26
2.1.4. Incidência de Raios e o Índice Ceráunico	28
2.1.5. Densidade de Descargas Atmosféricas	30
2.2. Efeitos das Descargas Atmosféricas no Sistema Elétrico	31
3 Os Dados de Descargas Atmosféricas	34
3.1. Atributos do Banco de Dados de Descargas Atmosféricas	35
3.2. Característica de Agrupamento dos Dados de Descargas Atmosféricas	37
3.3. Dados de Desligamentos de Linhas de Transmissão	43
4 Modelo de Determinação em Tempo Real do Risco de Desligamentos em Linhas de Transmissão por Raios – RDLR	46
4.1. Pré-processamento dos Dados	47
4.1.1. Definição do Espaço Amostral	47
4.1.2. Agrupamento dos Dados de Descargas Atmosféricas das Cenas	48
4.1.3. Seleção dos Agrupamentos de Descargas Atmosféricas Relevantes	50
4.1.4. Seleção das Variáveis de Entrada por Análise de Componentes Principais	50
4.2. O Modelo da Rede Neural MLP	53
4.2.1. Variáveis de Entrada	53
4.2.2. A Camada Intermediária	54
4.2.3. As Variáveis de Saída	55

4.3. Definição da Determinação do Risco em Tempo Real do Desligamento em Linhas de Transmissão por Descargas Atmosféricas	56
4.4. Modelo RDLR Dinâmico (Tempo Real)	57
5 Estudo de Caso	59
5.1. Análise dos Dados de Desligamento de Linhas de Transmissão	59
5.2. Pré-processamento dos dados de desligamento de linhas de transmissão e descargas atmosféricas	61
5.2.1. Critério de Seleção das Linhas de Transmissão para Estudo	61
5.2.2. Critério para Definição do Espaço Amostral de Descargas Atmosféricas	66
5.3. Agrupamento do espaço amostral de descargas atmosféricas	69
5.4. O Modelo da Rede Neural <i>Multilayer Perceptron</i> (MLP)	75
5.4.1. Variáveis de Entrada dos Agrupamentos de Descargas Atmosféricas	75
5.4.2. Seleção dos Agrupamentos Relevantes	76
5.4.3. Seleção das Variáveis de Entrada	83
5.4.4. Definição da Camada Intermediária da Rede Neural MLP	92
5.4.5. As Variáveis de Saída da Rede Neural MLP	93
5.4.6. A Determinação do Risco no Modelo RDLR	94
5.5. Análise dos Resultados	96
6 Conclusões e Recomendações	106
7 Referências Bibliográficas	108
ANEXO I. Aspectos Teóricos sobre Agrupamento	112
A. Agrupamento dos Dados	113
B. Aplicações	115
C. O Problema do Agrupamento	116
D. Etapas do Processo de Agrupamento	117
ANEXO II. Avaliação da Tendência de Agrupamento dos Dados de Descargas Atmosféricas	119

E. Tendência de Agrupamento	120
ANEXO III. Agrupamento dos Dados de Descargas Atmosféricas	124
F. Metodologia de Agrupamento	125
G. Agrupamento Baseado em Densidade e DBSCAN	129
H. Agrupamento dos Dados de Descargas Atmosféricas Baseado em Densidade	136
I. Agrupamento do espaço amostral de descargas atmosféricas - Estudo de Caso	155

Lista de Figuras

Figura 2.1 – Nuvem cumulonimbus	23
Figura 2.2 – Etapas de um relâmpago nuvem-solo negativo	25
Figura 2.3 - Eficiência da Rede de Detecção de Descargas Atmosféricas	28
Figura 2.4 – Mapa isoceráunico do Brasil	29
Figura 2.5 – Mapa de densidade de descargas atmosféricas	31
Figura 2.6 – Descarga direta em um condutor fase de uma linha de transmissão 500 kV	32
Figura 3.1 – Curva típica de descarga atmosférica	35
Figura 3.2 – Amostra randomicamente distribuída no espaço	38
Figura 3.3 – Estatística de Hopkins para amostra randomicamente distribuída no espaço ($H \approx 0,5$)	39
Figura 3.4 – Amostra uniformemente distribuída no espaço	39
Figura 3.5 – Estatística de Hopkins para amostra uniformemente distribuída no espaço ($H \approx 0$)	40
Figura 3.6 – Amostra com dados agrupáveis	41
Figura 3.7 – Estatística de Hopkins para amostra com dados agrupáveis ($H \approx 1$)	41
Figura 3.8 – Amostra de descargas atmosféricas dia 20/09/2014	42
Figura 3.9 – Estatística de Hopkins (descargas atmosféricas 20/9/2014)	42
Figura 4.1 – Modelo RDLR Conceitual – Otimização dos Parâmetros	47
Figura 4.2 - A Rede Neural MLP do Modelo RDLR	56
Figura 4.3 – Modelo RDLR Dinâmico (Tempo Real)	58
Figura 5.1 - Principais causas de desligamentos de linhas de transmissão – 2002 a 2015	61
Figura 5.2 - Espaço amostral geográfico de descargas atmosféricas em torno da LT 440 kV TAQUARUÇU/JUPIA associados a desligamentos da linha de transmissão	67
Figura 5.3 - DBSCAN para $Eps=0,05$ e $MinPts=5$ (Intervalo 02:00:04 a 02:30:04)	71
Figura 5.4 – DBSCAN para $Eps=0,33$ e $MinPts=5$ (Intervalo 00:00:00 a 00:30:00)	72
Figura 5.5 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 3	72

Figura 5.6 - Agrupamento das descargas atmosféricas sem a ocorrência de desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 11	73
Figura 5.7 - Agrupamento das descargas atmosféricas sem a ocorrência de desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 6	73
Figura 5.8 - Agrupamento das descargas atmosféricas sem a ocorrência de desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 29	74
Figura 5.9 – Matriz de Confusão – Agrupamentos ordenados pela distância do centroide do grupo à LT	77
Figura 5.10 – Desempenho da Rede Neural – Agrupamentos ordenados pela distância do centroide do grupo à LT	78
Figura 5.11 – Matriz de Confusão – Agrupamentos ordenados pela quantidade de raios	79
Figura 5.12 – Desempenho da Rede Neural – Clusters ordenados pela quantidade de raios	80
Figura 5.13 – Matriz de Confusão – Agrupamentos ordenados pela razão da quantidade de raios e a distância do centroide à LT	81
Figura 5.14 – Desempenho da Rede Neural – Agrupamentos ordenados pela razão da quantidade de raios e distância do centroide à LT	82
Figura 5.15 – Gráfico de barras das Componentes Principais x Variância	84
Figura 5.16 – Score das entradas sobre as duas primeiras componentes principais	85
Figura 5.17 - Score das entradas sobre as três primeiras componentes principais	85
Figura 5.18 - Análise das variáveis de entrada - 3 Componentes Principais	88
Figura 5.19 - Análise das variáveis de entrada - 8 Componentes Principais	89
Figura 5.20 - Análise das variáveis de entrada - 30 Componentes Principais	92
Figura 5.21 - A Rede Neural MLP do Modelo RDLR	94
Figura 5.22 - Modelo RDLR para determinação do risco de desligamentos de linhas de transmissão por descargas atmosféricas.	95
Figura 5.23 - Matrizes de Confusão sobre o conjunto de teste das 25 configurações de Rede Neural MLP e melhor instância	99
Figura 5.24 - Comportamento do erro médio das 50 instâncias sobre o conjunto de teste	101
Figura 5.25 - Comportamento do erro médio das 50 instâncias sobre o conjunto total	101
Figura 5.26 - Comportamento do erro médio da melhor instância sobre o	

conjunto de teste	103
Figura 5.27 - Comportamento do erro médio da melhor instância sobre o conjunto total	103
Figura 5.28 - Erro médio dos conjuntos de teste e total (44 variáveis de entrada e 8 neurônios na camada intermediária)	105
Figura 5.29 – Detalhe do erro médio dos conjuntos de teste e total (44 variáveis de entrada e 8 neurônios na camada intermediária)	105

Lista de Tabelas

Tabela 3.1 – Atributos dos dados de descargas atmosféricas	37
Tabela 3.2 – Códigos das causas dos desligamentos das linhas de transmissão	43
Tabela 4.2 – Descrição das Variáveis de Entrada	53
Tabela 4.3 – Descrição das Variáveis de Saída	55
Tabela 5.1 - Dados sobre desligamentos de linhas de transmissão e quantidade de descargas atmosféricas	60
Tabela 5.2 - Número de descargas atmosféricas por nível de tensão, no critério definido	62
Tabela 5.3 - Quantidade de desligamentos no setor 500 kV	63
Tabela 5.4 - Quantidade de desligamentos no setor 440 kV	64
Tabela 5.5 – Critério de Seleção da Linha de Transmissão	65
Tabela 5.6 – Cenas sem desligamentos da LT 440 kV TAQUARUÇU/JUPIÁ ordenadas pela quantidade de raios	68
Tabela 5.7 – Resultados do agrupamento com algoritmos DBSCAN	70
Tabela 5.8 – Valores dos parâmetros Eps e MinPts analisados no estudo	70
Tabela 5.9 – Análise comparativa dos métodos de seleção dos agrupamentos	82
Tabela 5.10 - Variância e Variância Acumulada	86
Tabela 5.11 - Indicação das variáveis de entrada mais relevantes - 3 CP	87
Tabela 5.12 - Indicação das variáveis de entrada mais relevantes - 8 CP	88
Tabela 5.13 - Indicação das variáveis de entrada mais relevantes - 30 CP	91
Tabela 5.14 - Erro médio das 50 instâncias do conjunto de teste	100
Tabela 5.15 - Erro médio das 50 instâncias do conjunto total	101
Tabela 5.16 - Erro médio da melhor instância do conjunto de teste	102
Tabela 5.17 - Erro médio da melhor instância do conjunto total	103
Tabela 5.80 - Erro médio dos conjuntos total e teste (44 variáveis de entrada e 8 neurônios na camada intermediária)	104

1

Introdução

1.1. Motivação

Descargas atmosféricas são fenômenos eletromagnéticos que ocorrem em grandes proporções, apresentando um comportamento absolutamente aleatório e imprevisível. Devido a sua grande extensão territorial, predominantemente em região tropical e sofrendo frentes oriundas do polo sul, o Brasil é um dos países com as maiores incidências de raios no mundo, chegando a 100 milhões por ano [1]. As descargas atmosféricas são de grande importância para o setor elétrico, sendo frequentemente responsáveis por desligamentos de linhas de transmissão, as quais podem desencadear uma sequência de eventos que levem o sistema elétrico interligado ao colapso. As longas extensões de linhas de transmissão determinam uma probabilidade significativa de incidência direta de descargas atmosféricas nestes equipamentos.

O caráter estratégico das linhas para o fornecimento de energia e a constatação de que descargas atmosféricas estão entre as principais causas de desligamentos, sugerem a necessidade de alertas que possibilitem ações operativas antecipadas para manter a segurança do sistema elétrico. Do ponto de vista econômico, as concessionárias transmissoras recebem pela disponibilidade dos seus ativos. Quando uma linha de transmissão está fora de serviço por algum motivo não programado, tal como por descargas atmosféricas, e período de indisponibilidade ultrapasse o padrão de duração de desligamento aceitável (valor limite), incidirá sobre a receita da linha a Parcela Variável por Indisponibilidade (PVI), com impacto relevante diretamente sobre a receita total da linha de transmissão [2].

Muitos estudos têm sido conduzidos com o propósito de desenvolver um sistema de monitoramento de descargas atmosféricas baseado em sensores, que forneçam maior eficiência e precisão de localização. Nesse sentido, um desses estudos [3] é realizado sobre um sistema de monitoramento de descargas atmosféricas para o Nordeste do País. Nesse estudo, foram analisados os

resultados de um sistema de monitoramento de raios, a partir da nova rede de detecção de descargas atmosféricas a longa distância, visando a auxiliar as operações da concessionária de energia elétrica local, bem como a um melhor planejamento da infraestrutura da rede.

Essa mesma abordagem é observada em trabalhos no âmbito internacional [4], onde é descrita a rede nacional de detecção de raios nos Estados Unidos (NLDL), fornecendo a localização e informações de descargas atmosféricas para beneficiarem as empresas de energia elétrica, que possuem seus ativos expostos ao grave problema de interrupções por descargas atmosféricas.

Outro trabalho aborda a estimativa de probabilidade de ocorrência de relâmpagos [5], que apresenta uma relação entre descargas elétricas e nuvens “penetrativas”, isto é, nuvens com convecção profunda alcançando a tropopausa. O sistema permite alertar em tempo real a possível ocorrência de descargas elétrica sobre o continente da América do Sul.

Entretanto, os estudos encontrados atualmente estão orientados para a eficiência das redes de detecção de descargas atmosféricas e na identificação de condições climáticas que identifiquem a ocorrência de raios de forma preditiva, sem uma correlação com as linhas de transmissão que possa gerar alertas de riscos de desligamentos dessas linhas de transmissão.

A inexistência de ferramentas que auxiliem a tomada de decisão na operação em tempo real do sistema elétrico, em função das perturbações causadas por descargas atmosféricas, motivou esse estudo para a proposição de uma abordagem inovadora para a determinação dos riscos de desligamentos de linhas de transmissão causadas por esse fenômeno.

1.2. Objetivos

Desenvolver um modelo para a determinação antecipada do risco de desligamentos de linhas de transmissão causados por raios, de forma a apoiar as decisões operativas em Sistemas Elétricos de Potência, mitigando os danos causados por esses distúrbios na operação do sistema.

1.3. Descrição do Trabalho

Grande parte do território brasileiro sofre grande incidência de descargas atmosféricas. Agregado ao fato de possuir um solo com características bem

desfavoráveis, tal como elevados valores de resistividade do solo, tornam a incidência de raios a causa da maioria dos curtos-circuitos e interrupções de serviço não programadas das linhas de transmissão no Brasil.

Diante desse cenário, o trabalho descreve um modelo para a determinação do risco de desligamento de linhas de transmissão por descargas atmosféricas, fornecendo informações de apoio a decisões antecipadas na operação de Sistemas Elétricos de Potência, tanto no Brasil quanto em outros países, minimizando os riscos de interrupção de energia elétrica para a sociedade.

Esse modelo, denominado por RDLR (Risco de Desligamento de Linhas de Transmissão por Raios), envolveu a elaboração de 3 etapas distintas, sendo a primeira definida pela coleta e pré-processamento dos dados, para a seleção do espaço amostral de descargas atmosféricas relacionadas a uma linha de transmissão. Na etapa seguinte, as descargas atmosféricas do espaço amostral definido na primeira etapa são agrupados e novamente selecionados, conforme sua relevância. Por fim, na última etapa é realizado um tratamento das variáveis de entrada redundantes através do método PCA - *Principal Component Analysis* para em seguida serem processados por uma Rede Neural *Multilayer Perceptron* – MLP, a qual identifica duas classes (MLP com duas saídas): a ocorrência ou não de desligamento da linha de transmissão devido às informações sobre as descargas atmosféricas. Esses resultados (as saídas da rede MLP) são então interpretados no módulo decisório, que determina o risco de desligamento de linhas de transmissão por descargas atmosféricas.

Esse estudo se baseou em um banco de dados com um grande volume de informações, reunindo 29262 registros sobre desligamentos de linhas de transmissão, associado a outro banco de dados de descargas atmosféricas, obtido através das redes de detecção de descargas atmosféricas, contendo 190 milhões de registros. A associação das informações contidas nestes dois bancos de dados possibilitou a constatação de que a ocorrência de descargas atmosféricas é uma das principais causas responsáveis pelos desligamentos de linhas de transmissão.

A partir dessas informações contendo dados de descargas atmosféricas e registros dos desligamentos de linhas de transmissão em todo o Brasil, com a utilização de um banco de dados espacial, foram propostos critérios para a definição do conjunto amostral de descargas atmosféricas a uma determinada distância das linhas de transmissão. Dessa forma, foi definido um período de observação baseado no intervalo de tempo anterior ao instante dos desligamentos das linhas.

A exploração dessas informações permitiu observar que as regiões do Brasil com pouca cobertura das redes de detecção de descargas atmosféricas prejudicam a análise de desligamento em linhas de transmissão localizadas nessas áreas, sendo necessário definir metodologias para a escolha das linhas localizadas em regiões com boa eficiência das redes de detecção de raios, permitindo o ajuste adequado do modelo proposto nesse trabalho.

Após a definição da linha de transmissão analisada nesse estudo, foi realizada a segmentação do espaço amostral, através de um algoritmo baseado em densidade (DBSCAN), para agrupar os dados de descargas atmosféricas, que possuem regiões densas, de formas não circulares e distintas.

O estudo definiu critérios para a determinação dos grupos mais relevantes da amostra, para cada desligamento associado à linha transmissão, a fim de melhorar o desempenho do modelo. Os critérios se basearam na distância do centroide do agrupamento à linha de transmissão, na quantidade de raios por grupo e um terceiro critério agregando a distância do centroide do agrupamento à linha de transmissão e quantidade de descargas nos grupos. Dessa forma, foram obtidos os melhores resultados, considerando esse último critério citado, sendo identificado um desempenho superior aos demais critérios avaliados nesse estudo.

Os grupos selecionados são então processados por uma Rede Neural *Multilayer Perceptron* – MLP, cuja capacidade de generalização e adaptabilidade são características importantes para sua utilização no modelo RDLR, além de possuir a característica de separar padrões que não são linearmente separáveis. Os atributos das descargas atmosféricas que compõem os agrupamentos foram analisados para a seleção das variáveis de entrada que representam o grupo. As saídas da rede neural representam as classes dos padrões de raios relacionados a desligamento de linhas de transmissão e padrões de descargas que não estão associados a padrões de desligamento desses equipamentos. O módulo decisório interpreta as classes através do maior valor de saída e determina o risco de desligamento das linhas de transmissão em função desse mesmo valor.

O modelo RDLR apresentou resultados bastante satisfatórios, diferenciando de forma correta cenários de descargas atmosféricas que de fato se relacionavam ao desligamento da linha de transmissão daqueles cenários que não causaram o desligamento desse equipamento. O modelo RDLR é capaz de informar o risco de desligamento de linhas de transmissão, se mostrando como uma inovadora e valiosa ferramenta para apoiar decisões antecipadas na operação dos Sistemas Elétricos de Potência.

1.4. Organização do Trabalho

A presente dissertação está organizada em seis capítulos, sendo cinco além deste.

O Capítulo 2 consiste em uma abordagem teórica dos desligamentos de linhas de transmissão por descargas atmosféricas, englobando conceitos sobre descargas atmosféricas, descrição dos sistemas de detecção desses eventos e o mapeamento da densidade de descargas atmosféricas no território brasileiro. Além disso, também são apresentados os efeitos das descargas atmosféricas no sistema elétrico.

O Capítulo 3 descreve as informações reais sobre descargas atmosféricas, fornecidas pelas redes de detecção, e os dados de desligamentos de linhas de transmissão em todo o território brasileiro. O banco de dados de descargas atmosféricas foi avaliado previamente para observar sua tendência de agrupamento dos dados. Para essa análise, foi utilizada a Estatística de Hopkins para identificar se o conjunto de dados de descargas atmosféricas possui tendência a formar grupos em suas localizações geográficas, avaliando a existência de estruturas não randômicas na amostra.

O Capítulo 4 descreve o modelo genérico RDLR – Risco de Desligamento de Linhas de Transmissão por Raios, proposto nesta dissertação para a determinação de risco de desligamento de linhas de transmissão.

O Capítulo 5 apresenta o Estudo de Caso, utilizando o modelo descrito no Capítulo 4. O Estudo de Caso aborda a análise dos dados de desligamento de linhas de transmissão e o pré-processamento dos dados para a segmentação do espaço amostral de descargas atmosféricas. Nesse capítulo, também são apresentados os resultados sobre os testes com as diferentes arquiteturas da Rede Neural MLP, para a definição daquela que produziu melhores resultados na determinação do risco de desligamento de linhas de transmissão para o espaço amostral definido no trabalho.

O Capítulo 6 contém as conclusões e recomendações geradas sobre este trabalho.

Por fim, no Anexo I são descritos aspectos teóricos sobre agrupamento. O Anexo II contém o registro completo da avaliação da tendência de agrupamento dos dados de descargas atmosféricas, através da Estatística de Hopkins. No Anexo III são apresentados os estudos com diversos parâmetros testados com o

método baseado em densidade DBSCAN e os agrupamentos do espaço amostral relacionados à linha de transmissão analisada nesse trabalho.

2

Desligamentos de Linhas de Transmissão por Descargas Atmosféricas

As descargas atmosféricas constituem-se no principal elemento responsável pelos desligamentos não programados dos sistemas elétricos de potência.

Quando uma descarga atmosférica atinge uma linha de transmissão ou sua vizinhança, uma sobretensão é desenvolvida através de seu isolamento. Se essa tensão exceder a sua suportabilidade, ocorrerá uma falha de isolamento. A consequência dessa falha é a formação de um arco elétrico, o qual será mantido pela tensão do sistema, tornando-se então necessária a operação dos dispositivos de proteção para a eliminação deste curto-circuito [6].

2.1.

O Fenômeno da Descarga Atmosférica

2.1.1.

Conceitos Básicos

Uma descarga atmosférica pode ser definida como uma descarga elétrica transitória de curta duração, com uma alta corrente associada, que usualmente atinge quilômetros de extensão [6]. Tal descarga ocorre quando nuvens em uma determinada região da atmosfera atingem uma quantidade de cargas elétricas de forma a gerar campos elétricos tão intensos que superam a rigidez dielétrica do ar, causando a sua ruptura. Esta suportabilidade é da ordem de 30 kV/cm para o ar seco nas condições atmosféricas padronizadas, mas pode ser bem menor, devido tanto à redução da pressão atmosférica como também devido à elevada umidade na atmosfera [7]. Para que tal processo possa acontecer, é necessário que as condições ambientais estejam favoráveis. As descargas atmosféricas são associadas a nuvens de tempestade do tipo cumulonimbus (ver Figura 2.1), que geram chuvas fortes, trovões, raios e, ocasionalmente granizos.



Figura 2.1 – Nuvem cumulonimbus

Fonte: CARN (2009)

Os processos de eletrificação das nuvens de tempestade ainda são objetos de muito estudo e discussão pela comunidade científica internacional, mas modelos computacionais e simulações de pequena escala em laboratórios têm mostrado que fatores termodinâmicos (como perfil de temperatura da atmosfera, disponibilidade de água líquida, umidade relativa), associados com a disponibilidade de núcleos de condensação na atmosfera e a formação de cristais de gelo dentro da nuvem, têm papel fundamental na geração dos grandes centros de cargas elétricas a partir dos quais têm origem as descargas atmosféricas. Devido à grande variabilidade desses fatores, os quais determinam diretamente as características elétricas das nuvens, os relâmpagos, por consequência, também apresentam uma grande variabilidade de suas características físicas [8].

2.1.2. Tipos de Descargas Atmosféricas

Denominamos os tipos de relâmpagos em função de sua origem e do ponto onde terminam. Eles podem ocorrer (1) de uma nuvem para o solo (chamado nuvem-solo ou NS); (2) do solo para a nuvem (chamado solo-nuvem ou SN); (3) dentro da nuvem (chamado intranuvem ou IN); (4) entre nuvens (NN); (5) de uma nuvem para a atmosfera acima dela; e, finalmente, (6) de uma nuvem para um ponto qualquer no ar (descargas no ar) [9].

Os relâmpagos IN e NN são os mais frequentes (cerca de 80% do número total global de relâmpagos), mas de difícil estudo por ocorrerem em meio à opacidade da nuvem.

Os relâmpagos que envolvem o solo, nuvem-solo e solo-nuvem, embora não sejam os mais frequentes, são os mais estudados por estarem mais acessíveis à observação visual, por conseguinte à observação instrumentada, e por causa do perigo que apresentam. Os relâmpagos que envolvem o solo são denominados usualmente de raios. Estes, por sua vez, podem ser divididos em dois tipos: os negativos e os positivos, dependendo da polaridade predominante das cargas na região da nuvem onde se originam. Assim, existem quatro subtipos: NS negativo, NS positivo, SN negativo e SN positivo [9].

Os relâmpagos nuvem-solo negativos se originam no centro negativo de cargas próximo à base da nuvem e, em geral, são mais frequentes que os positivos. Os relâmpagos positivos são menos conhecidos, e só foram estudados mais detalhadamente nas últimas décadas. A polaridade da descarga é determinada, portanto, pela polaridade da carga (resultante) depositada no solo.

Os relâmpagos solo-nuvem são raros em comparação com os nuvem-solo, originando-se geralmente de pontos elevados tais como torres, edifícios, montanhas e árvores.

Um relâmpago nuvem-solo negativo segue geralmente nove etapas, conforme Figura 2.2, iniciando com uma descarga entre o centro positivo menor na base da nuvem e o centro negativo para a liberação dos elétrons ligados às partículas de água ou gelo. Uma fraca descarga luminosa, geralmente não visível, se propaga em direção ao solo, criando um canal ionizado (caminho condutor). Essa descarga, denominada líder escalonado, propaga-se para baixo até aproximadamente 5 a 50 m do solo, com velocidade média da ordem de 10⁵ m/s. O líder escalonado geralmente ramifica-se, mas nem todos os seus ramos estendem-se até o solo. Uma descarga conectante surge com a intensificação do campo elétrico entre a parte frontal do líder escalonado e o solo. Tal descarga geralmente parte de objetos pontiagudos existentes na superfície, como árvores e edifícios, e desloca-se em direção à parte frontal do líder escalonado [9].

Quando o encontro ocorre, a descarga conectante transforma-se numa descarga forte e de intensa luminosidade, denominada descarga de retorno, a qual é vista como um clarão. Ainda que, preferencialmente, o relâmpago atinja os objetos mais altos, isso nem sempre ocorre. Na descarga de retorno, as cargas negativas (elétrons) presentes no canal já ionizado pelo líder escalonado são aceleradas, o que cria uma corrente ascendente (os elétrons fluem da nuvem para o solo) com picos de 30 kA. A velocidade de propagação da frente de onda (ou da luminosidade ao longo do canal) é da ordem de 10⁸ m/s (um terço da velocidade da luz). A duração de tal descarga é de aproximadamente 70 microssegundos. O

líder contínuo, que é a ocorrência de outra descarga que percorre o canal do raio que está fracamente ionizado, deve ionizá-lo novamente. Caracteriza-se como um líder por possuir baixa intensidade de corrente. Possui velocidade característica de $(1-2) \times 10^7$ m/s. Esta etapa ocorre somente quando, de alguma forma, existem cargas disponíveis próximas ao canal criado anteriormente [9].

Após o líder contínuo ter ionizado o canal, outra descarga ocorre, sendo esta chamada descarga subsequente. A diferença essencial entre descargas de retorno e subsequente é que a primeira é precedida de líder escalonado, pois não existe canal algum, enquanto que a segunda é precedida por um líder mais veloz, que apenas refaz um canal previamente existente [9].

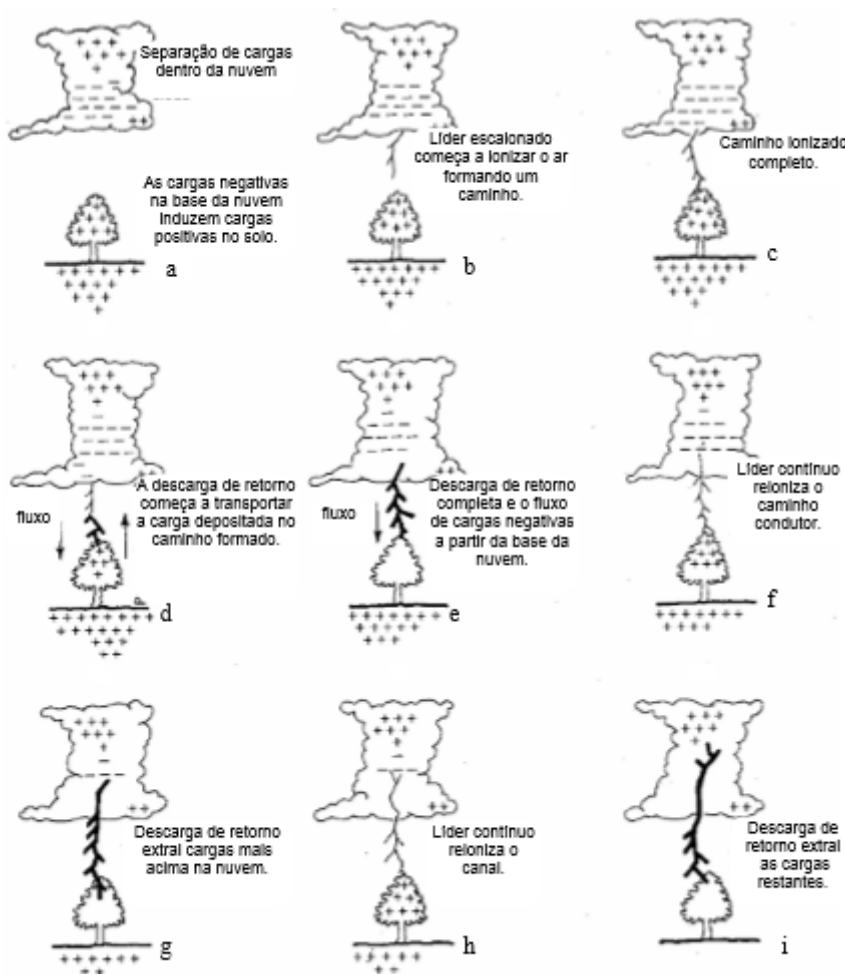


Figura 2.2 – Etapas de um relâmpago nuvem-solo negativo

Fonte: modificada de Lutgens e Tarbuck (1982, p. 244)

Os relâmpagos nuvem-solo positivos seguem, de um modo geral, as mesmas etapas descritas para os negativos, porém com algumas diferenças. Em

geral, iniciam-se a partir de um líder com uma luminosidade mais fraca comparada ao líder escalonado de um NS negativo. Tal líder propaga-se a partir de uma região de cargas positivas dentro da nuvem, não apresentando etapas e sim uma luminosidade contínua, porém com variações periódicas de intensidade. Outra característica importante é o número de descargas de retorno: somente uma na maioria das vezes. Costumam também apresentar corrente contínua mais longa (transferem maior quantidade de carga elétrica para o solo) do que os NS negativos, tornando-se mais destrutivos e sendo responsáveis pela maioria dos incêndios [10].

2.1.3.

Sistema de Detecção e Localização de Descargas Atmosféricas

Um sistema de detecção de descargas atmosféricas consiste em uma rede de sensores de radiação eletromagnética (EM) interligados a uma central de processamento capaz de detectar a ocorrência de descargas atmosféricas e identificar a localização (na superfície da Terra ou na atmosfera) de sua fonte de radiação [11].

Atualmente, existem diferentes tecnologias de detecção que permitem identificar e localizar tanto descargas nuvem-solo (NS) quanto nuvem-nuvem (NN) em função, basicamente, da faixa de frequência em que os sensores operam. No caso das descargas NS, a localização da fonte de radiação tende a coincidir com o ponto de impacto do raio na superfície, com um erro intrínseco devido ao próprio processo de detecção e localização. O erro no posicionamento, ou mais comumente denominado de precisão de localização (PL), é resultado de uma soma de erros sistemáticos e randômicos, tanto nas medidas fornecidas pelos sensores quanto no algoritmo matemático de triangulação usado para calcular as coordenadas geográficas do evento. A precisão de localização também depende da faixa de frequência de operação dos sensores, a qual determina fundamentalmente o alcance de detecção da radiação eletromagnética e o tipo de descarga envolvida.

Além do erro de localização, outro parâmetro importante de uma rede é sua eficiência de detecção (ED), ou seja, qual o percentual de eventos que ela consegue registrar em relação ao número real de ocorrências, já que não existe um sistema de detecção que seja capaz de registrar a ocorrência de todos os eventos. Para a faixa de frequência baixa denominada *Low Frequency* - LF (que vai de 10 a 350 kHz), o limite da tecnologia atual estabelece uma precisão de

localização da ordem de 500 m e eficiência de detecção em torno de 90%, o que é bastante razoável se comparado com o alcance de uma rede, que pode chegar até 1000 km. Entretanto, a radiação na faixa de frequência de LF se propaga pela superfície da Terra (próximo ao solo), sofrendo atenuações diferenciadas em função dos obstáculos que encontra em seu caminho, conhecido por efeito de propagação, como florestas, cadeias de montanhas, cidades, lagos e até o próprio oceano. Por outro lado, as distorções sofridas pelo sinal não são suficientes para impedir sua correta identificação e, conseqüentemente, a detecção por parte dos sensores. Em função disso, o que determina realmente a qualidade da informação fornecida por uma rede de detecção é a linha de base entre os sensores, ou seja, a máxima distância de separação entre dois sensores consecutivos que minimiza os efeitos de propagação. Neste caso, para uma rede que opera na faixa de LF, considerando a precisão de localização PL~500 m e a eficiência de detecção ED~90%, a linha de base deve ser inferior a 350 km [11]. A Figura 2.3 representa o mapa de eficiência da Rede Integrada de Detecção de Descargas Atmosféricas – RINDAT.

As redes de detecção de descargas atmosféricas estão presentes em 120 países do mundo e geram informações usadas por mais de 5 mil instituições. Esses dados alimentam serviços de meteorologia, diversos setores da engenharia e atendem a demandas do setor elétrico e agropecuário. Também ajudam a salvar vidas ao possibilitar previsão e monitoramento dos raios. Estima-se que os dados gerados por esses sistemas reduzam em até 20% os prejuízos causados pelas descargas atmosféricas. No Brasil, o ônus dos raios chega a R\$ 1 bilhão por ano [11].

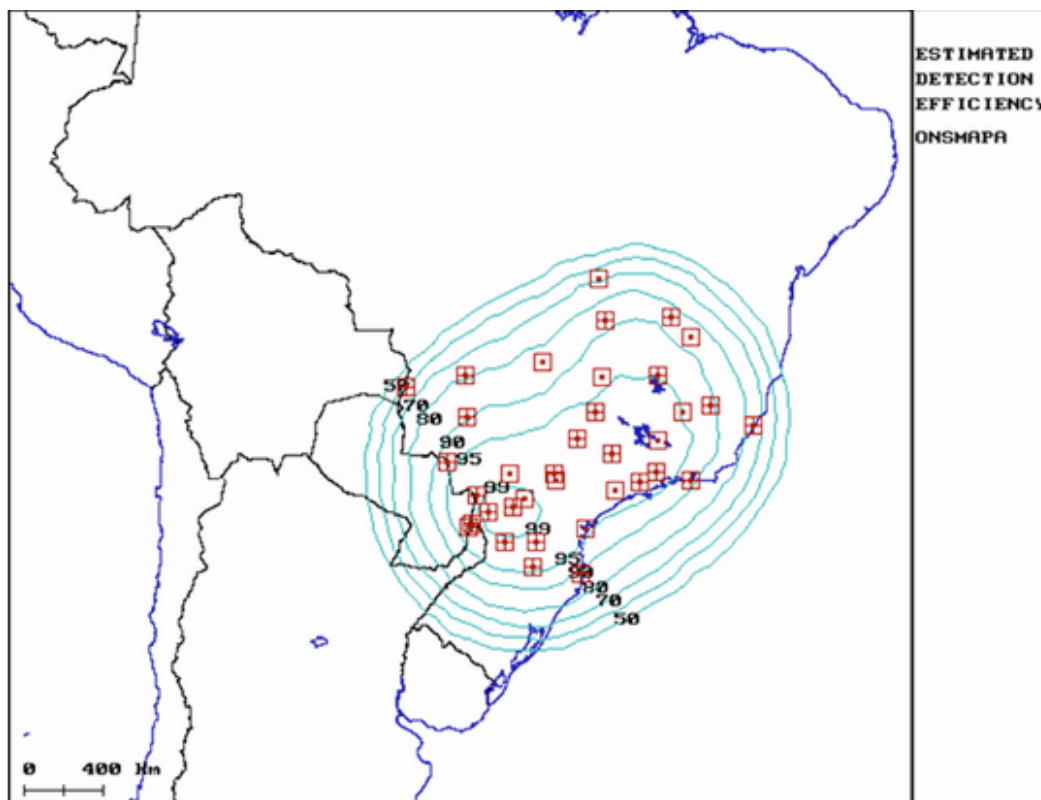


Figura 2.3 - Eficiência da Rede de Detecção de Descargas Atmosféricas

2.1.4. Incidência de Raios e o Índice Ceráunico

Um raio tende naturalmente a cair sobre os pontos mais elevados do relevo, como regiões montanhosas. Em um local qualquer a tendência é que as descargas incidam sobre árvores isoladas, antenas, chaminés e edificações mais elevadas [12]. Isso se deve principalmente ao fenômeno do poder das pontas que ocorre porque em um condutor eletrizado a carga tende a se acumular nas regiões pontiagudas. Cria-se um campo elétrico maior que nas regiões mais planas [13].

Quanto à natureza dos terrenos, estudos estatísticos demonstram que a incidência maior de descargas se dá em terrenos maus condutores, de formação granítica ou xistosa, em vez de em terrenos bons condutores como os calcários e solos de aluvião [14]. Isso ocorre porque, nos terrenos isolantes, a possibilidade de escoamento das cargas elétricas é menor, o que ajuda a manter e elevar o gradiente de tensão entre a nuvem e a terra. Em regiões onde há muita precipitação com tempestades a incidência de raios também é maior [12].

Dentre os parâmetros para estudos de desempenho de sistemas elétricos, a densidade de descargas para a terra é um dos parâmetros dos mais importantes

porque determina a quantidade de descargas incidentes a partir do número de dias com ocorrência de trovoadas por ano, que é o nível ceráunico da região.

Ao se unir em um mapa as localidades de um mesmo nível ceráunico, são obtidas curvas isoceráunicas, tal como se mostra para todo o Brasil na Figura 2.4, obtida da Norma Nacional NBR no 5419 (ABNT, 1993). O valor dos índices mínimo e máximo observados são, respectivamente, 5 e 140. Santa Catarina, por exemplo, possui um índice entre 40 e 60 trovoadas por ano, e Florianópolis um de índice ceráunico de 54.

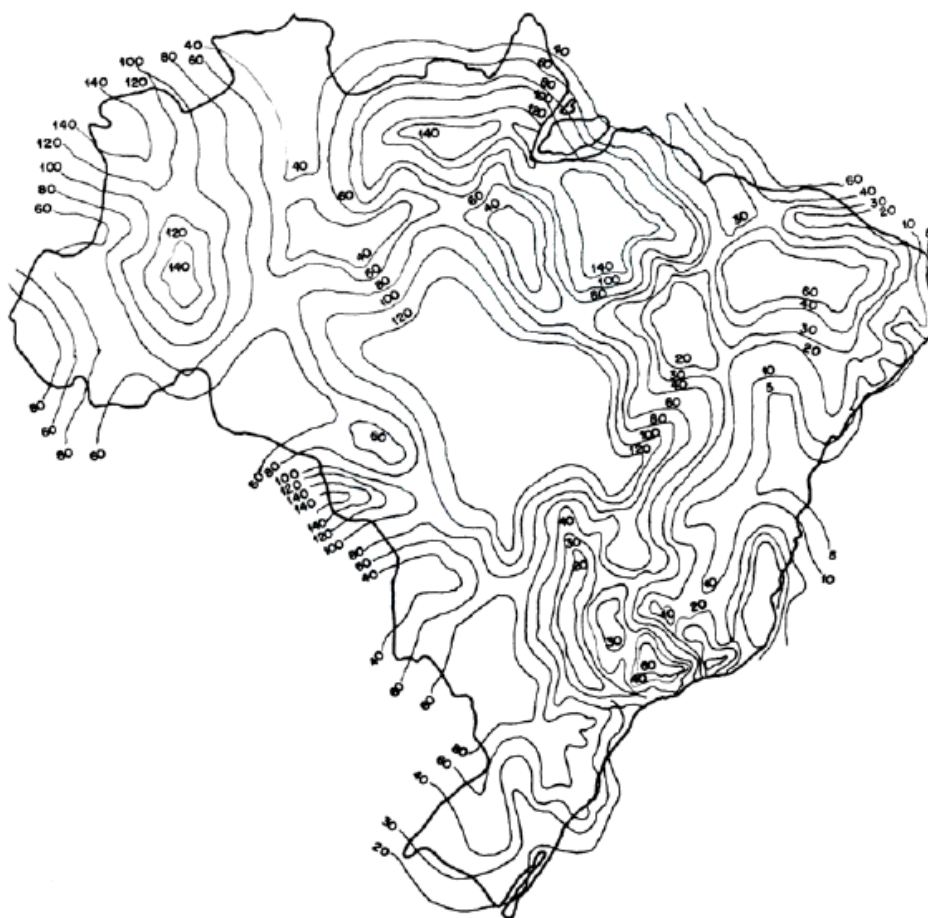


Figura 2.4 – Mapa isoceráunico do Brasil

2.1.5. Densidade de Descargas Atmosféricas

A frequência de descargas nuvem-solo (NS) é um parâmetro de grande significado para o desempenho do sistema elétrico. O parâmetro que traduz esse valor é a densidade de descargas atmosféricas para a terra. Esse índice estabelece uma medida de densidade por unidade de área, sendo expresso em número de descargas/km²/ano [14].

Os mapas de densidade de descargas para a terra, conforme apresentado na Figura 2.5, são usados para projetos de linhas de transmissão, estimando a quantidade de descargas atmosféricas para solo. A confiabilidade de uma linha de transmissão é dependente da sua exposição a descargas atmosféricas. Para determinar a exposição, o projetista necessita conhecer o número anual de descargas por unidade de área.

A densidade de descargas para a terra, N_g , pode ser estimada do nível ceráunico usando a equação de Anderson [14]:

$$N_g = 0,04 T_d^{1,25} [\text{descargas/km}^2/\text{ano}] \quad (1)$$

Onde T_d é o número de dias de trovoadas por ano (nível ceráunico).

A densidade de descargas para a terra também pode ser obtida a partir de dados de rede de detecção de descargas ou do contador de descargas. Se os dados estão presentes durante anos de coletas, isso tem uma vantagem na identificação das variações regionais.

As linhas de transmissão podem percorrer centenas de quilômetros e atravessar vários tipos de relevo e obstáculos no seu percurso, como vales, montanhas e rios. Dessa forma, uma linha pode percorrer regiões com densidade de descargas atmosféricas diferentes, principalmente devido à influência do relevo e da extensão da linha. A variação do relevo no percurso de uma linha pode afetar o seu desempenho pelo aumento da sua exposição à incidência de descargas. Uma torre de linha de transmissão localizada em um ponto elevado em relação às redondezas se constitui em um ponto preferencial de incidência de descargas.

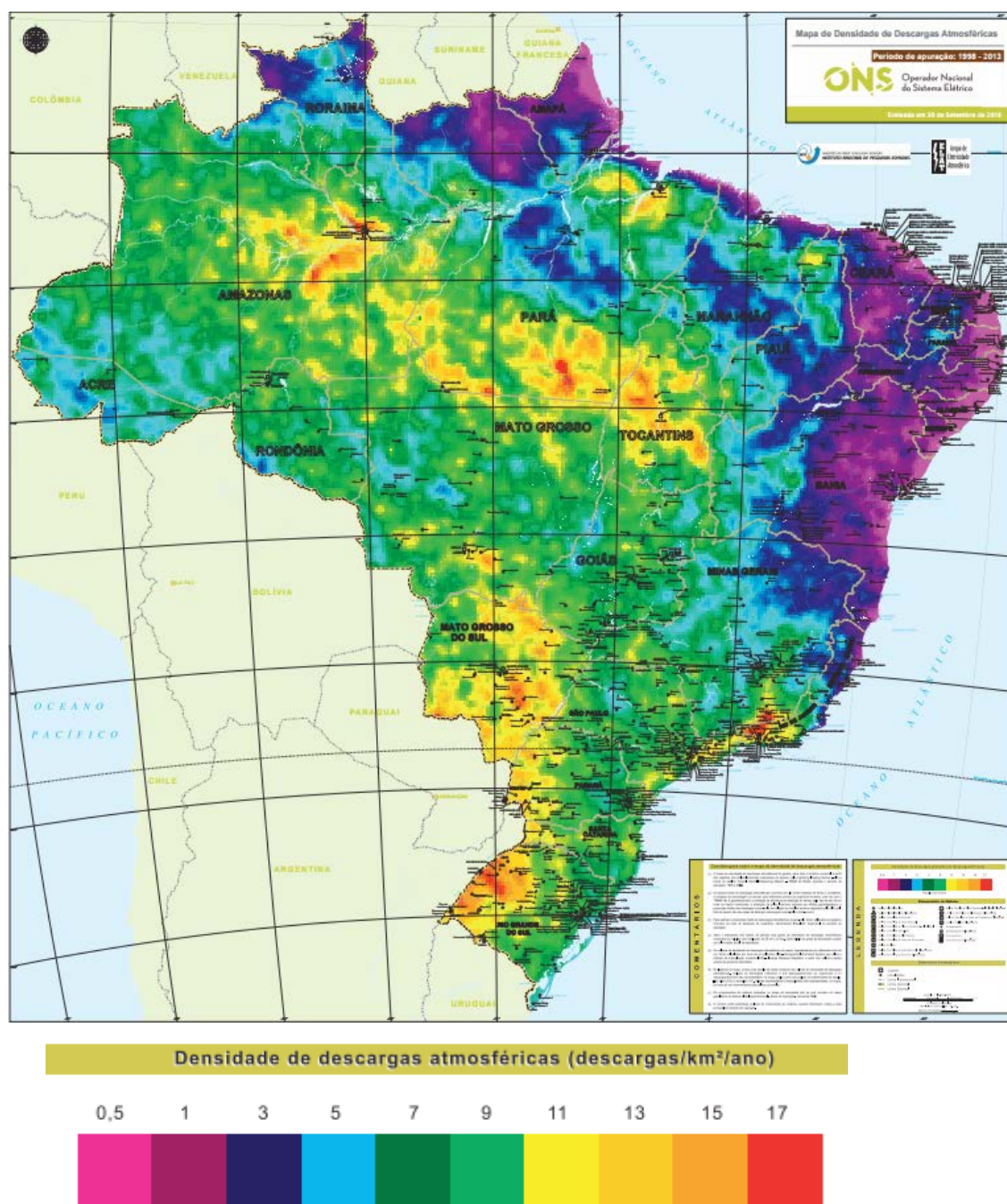


Figura 2.5 – Mapa de densidade de descargas atmosféricas

2.2.

Efeitos das Descargas Atmosféricas no Sistema Elétrico

A importância da incidência das descargas atmosféricas é bem caracterizada ao se notar que as mesmas são tipicamente a causa da maioria dos curtos-circuitos e interrupções de serviço não programadas das linhas de transmissão instaladas em várias partes do mundo. No Brasil esse problema é

especialmente importante, em virtude de em boa parte do território nacional haver grande incidência de descargas atmosféricas e o solo apresentar características bem desfavoráveis (elevados valores de resistividade do solo).

Existem muitos fatores que condicionam a frequência e a severidade da incidência de descargas atmosféricas numa determinada região e, conseqüentemente, do desempenho das linhas de transmissão nela instalada. Nota-se, por exemplo, a influência do clima, da orografia, da forma em que a proteção da linha foi instalada, do nível de isolamento da linha, entre vários outros itens.

O efeito das descargas em linhas de transmissão pode ser direto ou indireto (ver Figura 2.6). Os efeitos indiretos acontecem quando o raio incide próximo à linha de transmissão. Neste caso, a tensão induzida só produzirá efeitos mais sérios em linhas de transmissão de tensão abaixo de 69kV, que é o caso típico de linhas de subtransmissão e de distribuição. Nas linhas com tensão mais alta o efeito da indução pode ser desconsiderado. A descarga direta é aquela que o raio incide diretamente na linha de transmissão, podendo ocorrer de dois modos: no vão da linha de transmissão e na torre. Essas descargas geralmente afetam o desempenho das linhas de transmissão, gerando sobretensões transitórias resultantes acima da suportabilidade da isolação das linhas [15].



Figura 2.6 – Descarga direta em um condutor fase de uma linha de transmissão 500 kV

O efeito de uma descarga atmosférica sobre uma linha de transmissão depende basicamente: do projeto e das características construtivas da linha; da impedância do sistema de aterramento para surtos; da amplitude e da forma de

onda da corrente de descarga; do ponto de incidência da descarga (se diretamente sobre a rede ou em suas proximidades); das características de isolamento das linhas; do grau de proteção oferecido pelos cabos para-raios (no caso de sua existência), etc. Todos esses fatores devem ser levados em consideração quando da análise do desempenho das linhas de transmissão frente às descargas atmosféricas [15].

Do ponto de vista econômico, mesmo um desligamento mínimo não-programado de uma linha de transmissão pode causar prejuízos grandes para o cliente, sobretudo quando se trata de cliente industrial, cuja produção depende do fornecimento ininterrupto da energia. Por outro lado, as concessionárias transmissoras recebem pela disponibilidade dos seus ativos. Tomando o caso de linhas de transmissão como exemplo, quando alguma linha está fora de serviço por algum motivo não programado, caso o período de indisponibilidade ultrapasse o padrão de duração de desligamento aceitável (valor limite), incidirá sobre a receita da linha a Parcela Variável por Indisponibilidade (PVI), que tem impacto direto na receita da linha de transmissão [2].

3

Os Dados de Descargas Atmosféricas

A ocorrência de tempestades no território brasileiro tem sido observada mais frequentemente, causando grandes impactos sociais e ambientais em diversas regiões no Brasil. Consequentemente, o volume de dados de descargas atmosféricas é muito grande, sendo importante a existência de ferramentas que possibilitem as análises dos especialistas. As redes de detecção de descargas atmosféricas no Brasil podem fornecer até 14 milhões de eventos por mês durante os períodos de chuvas intensas, sendo necessário processar e analisar todas essas informações. Esses eventos são detectados pelas redes RINDAT (Rede Integrada Nacional de Descargas Atmosféricas) [17] e BRASILDAT (Sistema Brasileiro de Detecção de Descargas Atmosféricas) [16] e armazenados em uma base de dados histórica, desde 2001, com cerca de 400 milhões de registros.

A RINDAT é uma rede de sensores especializados com centrais de processamento que permitem detectar, em tempo real, as descargas atmosféricas, incluindo descargas tipo nuvem-solo, que é a maior parte das descargas que atingem o solo no território brasileiro. Essa rede de sensores é uma integração das redes anteriormente existentes no Brasil: a RINDAT é um convênio de cooperação técnico-científico entre as instituições que possuem sensores de detecção de descargas atmosféricas e centrais de processamento de informações desses sensores no Brasil, formada pela Companhia Energética de Minas Gerais (CEMIG), o Instituto Tecnológico SIMPAR, Eletrobrás FURNAS, o Instituto Nacional de Pesquisas Espaciais (INPE) e a rede de sensores pertencentes ao Sistema de Informações Integradas baseados no Sistema de Detecção de Descargas Atmosféricas (SIDDEM) [17]. O sistema registra o instante preciso de ocorrência, a localização do ponto de impacto e características físicas como intensidade e polaridade. Isso é realizado mediante um conjunto de sensores remotos, que detecta a radiação eletromagnética emitida pelos raios, na faixa entre 100 kHz a 300 kHz. Atualmente, a RINDAT possui 35 sensores. Os sinais recebidos são analisados, através de algoritmos específicos, e são eliminados aqueles cujas fontes não são descargas atmosféricas. Cada sensor que tenha detectado um evento válido envia os dados obtidos para um sistema central de processamento, que calcula os parâmetros de cada descarga.

A BRASILDAT é uma rede de detecção e monitoramento de descargas atmosféricas, operada pelo Grupo de Eletricidade Atmosférica (ELAT) do INPE, em São José dos Campos, e foi criada em 2011 [16]. Possui 70 sensores espalhados nas regiões Sul, Sudeste, Centro-Oeste e Nordeste do Brasil. Essa rede também registra instante e localização do ponto de impacto e características físicas, como intensidade e polaridade. Os dados são obtidos por meio de sensores que detectam a radiação eletromagnética emitida pelas descargas na faixa entre 10 Hz e 10 MHz. Os sinais são analisados e os ruídos são eliminados nesse processo.

3.1.

Atributos do Banco de Dados de Descargas Atmosféricas

Os sinais eletromagnéticos detectados e digitalizados pelos sensores de descargas atmosféricas são coletados pelas centrais de processamento que recebem e armazenam os dados brutos; realizam o cálculo das posições válidas das descargas atmosféricas detectadas através de algoritmos de triangulação e disponibilizam as informações de cada descarga, tal como data e hora com precisão de nanosegundo; localização (latitude e longitude) mais provável da incidência, estimativa do pico de corrente (kA), tempo de ascensão do pico de corrente (“risetime” em microssegundos), taxa de crescimento (“rate-of-rise” em kA/s), indicação da precisão da localização, através de uma elipse, assim como outros parâmetros. A Figura 3.1 representa o campo elétrico produzido por uma descarga atmosférica típica, detectada pelos sensores da rede.

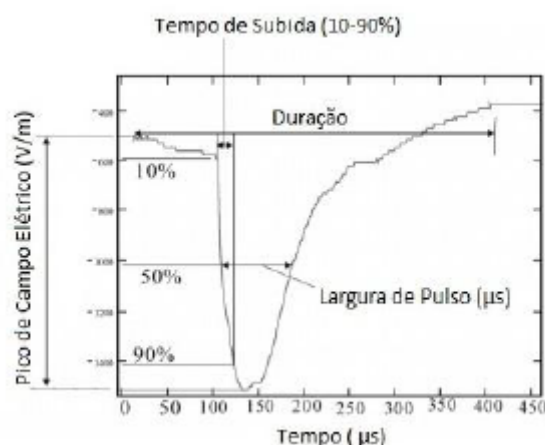


Figura 3.1 – Curva típica de descarga atmosférica

Fonte: Adaptado de Miki et al (2006)

A Tabela 3.1 contém os parâmetros calculados pelas centrais de processamento das redes de detecção para cada descarga atmosférica.

Tabela 3.1 – Atributos dos dados de descargas atmosféricas

Ano, mês e dia (GMT)
Hora, minuto, segundo, nanosegundo
Latitude e longitude (grau)
Pico de corrente estimado (kA)
Multiplicidade
Número de sensores utilizados na detecção da descarga
Graus de liberdade quando otimizando localização
Ângulo da elipse de incerteza (sentido horário)
Comprimento do semieixo maior da elipse (km)
Comprimento do semieixo menor da elipse (km)
Valor do chi-quadrado da otimização da localização
Tempo de ascensão (μs)
Tempo do pico até zero (μs)
Taxa máxima de crescimento (kA/s)
Indicador de nuvem (nuvem-nuvem ou nuvem-solo)
Indicador de ângulo (1 se informação de ângulo do sensor foi usada para cálculo da posição e 0 caso contrário)
Indicador de sinal (1 se informação de sinal é usada para cálculo da posição e 0 caso contrário)
Indicador de tempo (1 se informação de tempo é usada para cálculo da posição e 0 caso contrário)

3.2.**Característica de Agrupamento dos Dados de Descargas Atmosféricas**

Uma importante etapa para a análise e classificação de grandes conjuntos de informações é avaliar primeiramente a tendência de agrupamento dos dados, ou seja, se uma determinada amostra de dados pode ser representada através de agrupamentos significativos. Se uma amostra apresenta grupos com boa coesão e boa separação pode-se concluir que os dados apresentam tendência de agrupamentos.

É possível utilizar uma abordagem estatística para avaliação da tendência de agrupamento dos dados, como a equação de Hopkins [19]. Nesta técnica, deve-se gerar p registros uniformemente distribuídos pelo espaço e utilizar p registros da base real, conforme equação:

$$H = \frac{\sum_{i=1}^p w_i}{\sum_{i=1}^p u_i + \sum_{i=1}^p w_i} \quad (2)$$

Sendo u_i as distâncias da vizinhança mais próxima dos dados gerados artificialmente, e w_i dos dados reais. Se u_i e w_i forem semelhantes (significa que o agrupamento está próximo dos dados gerados artificialmente), H deve ser próximo de 0,5. Valores de H próximos de 1 indicam que os dados estão devidamente agrupados, e próximos de 0 que os dados estão uniformemente distribuídos pelo espaço. Essa equação mede a diferença entre os objetos do conjunto e objetos aleatórios, ou seja, mede a aleatoriedade dos objetos.

O banco de dados de descargas atmosféricas possui um grande volume de dados (400 milhões de registros), o que caracteriza um “big-data”, sendo importante avaliar previamente a tendência de agrupamento dos dados. Nesta análise foi utilizada a Estatística de Hopkins [19] para identificar se o conjunto de dados de descargas atmosféricas possui tendência a formar grupos em suas localizações geográficas, avaliando a existência de estruturas não randômicas na amostra. De acordo com a Estatística de Hopkins, amostras randomicamente distribuídas no espaço produzem resultados próximos de $H=0,5$.

A Figura 3.2 representa um exemplo de amostra randomicamente distribuída no espaço. O eixo x do gráfico apresentado na Figura 3.3 representa o valor de H , onde nessa amostra se apresenta em torno de 0,5.

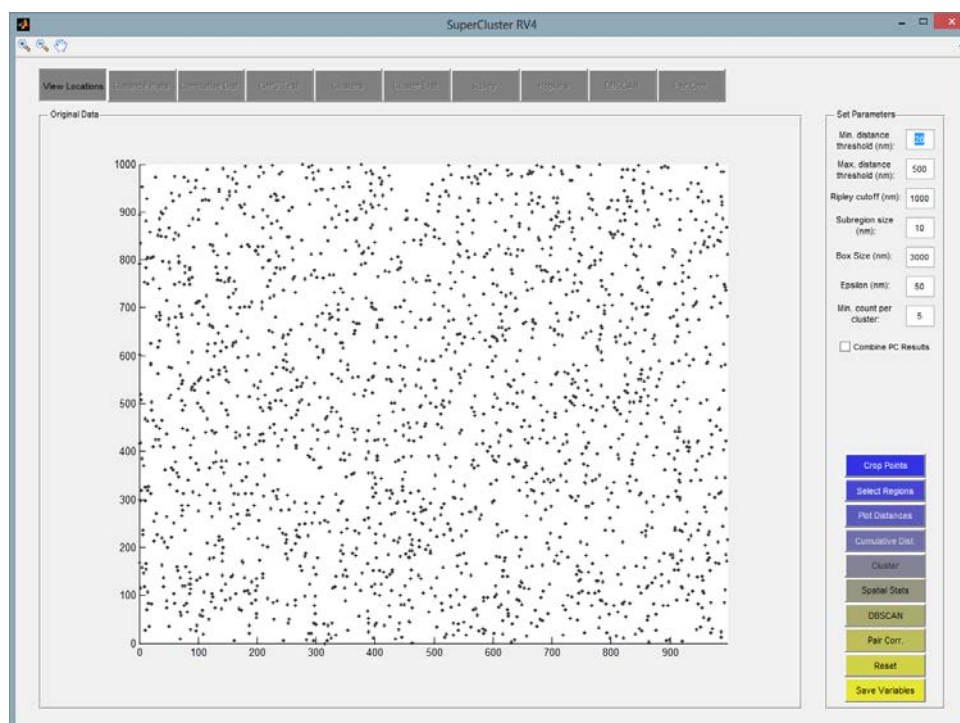


Figura 3.2 – Amostra randomicamente distribuída no espaço

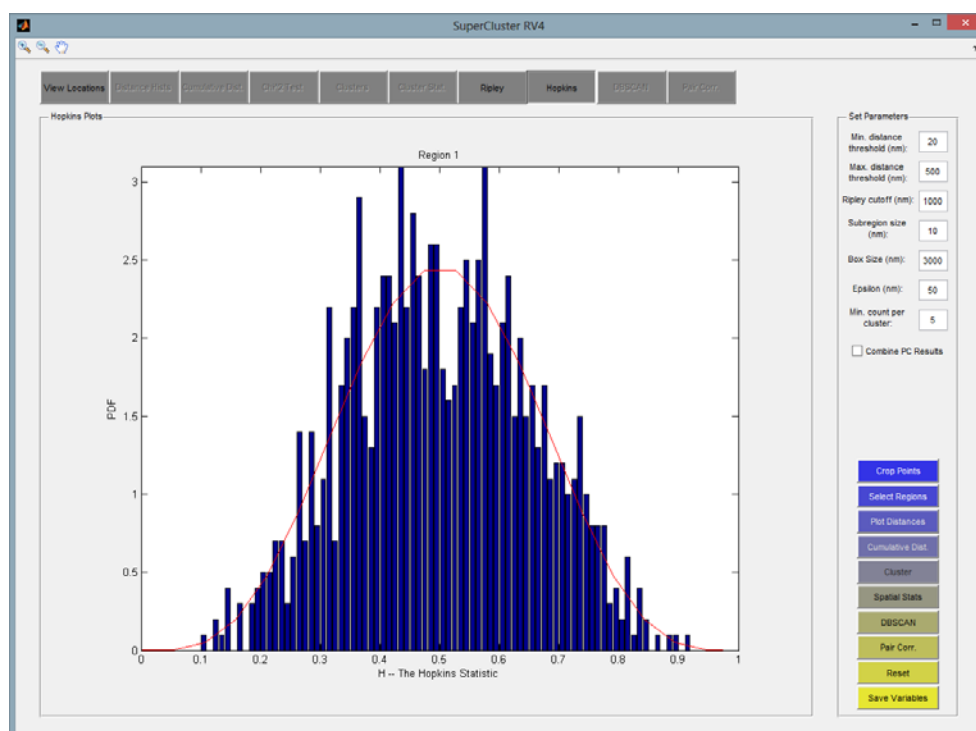


Figura 3.3 – Estatística de Hopkins para amostra randomicamente distribuída no espaço ($H \approx 0,5$)

Para amostras uniformemente distribuídas no espaço, conforme Figura 3.4, o resultado da Estatística de Hopkins é próximo a zero. Na Figura 3.5 é possível observar o valor de H no eixo x, próximo a zero.

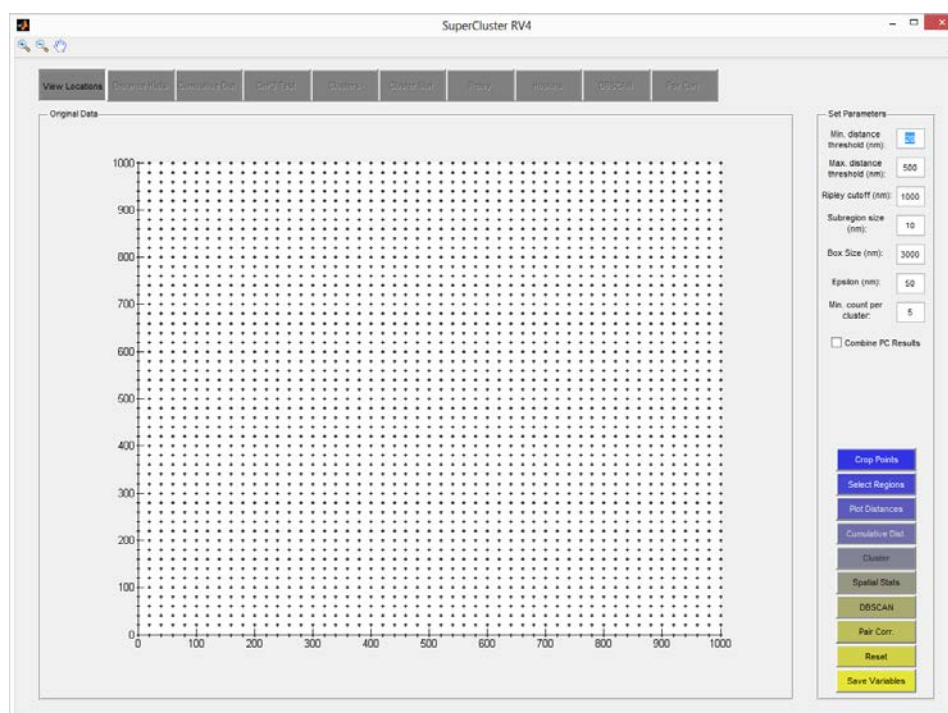


Figura 3.4 – Amostra uniformemente distribuída no espaço

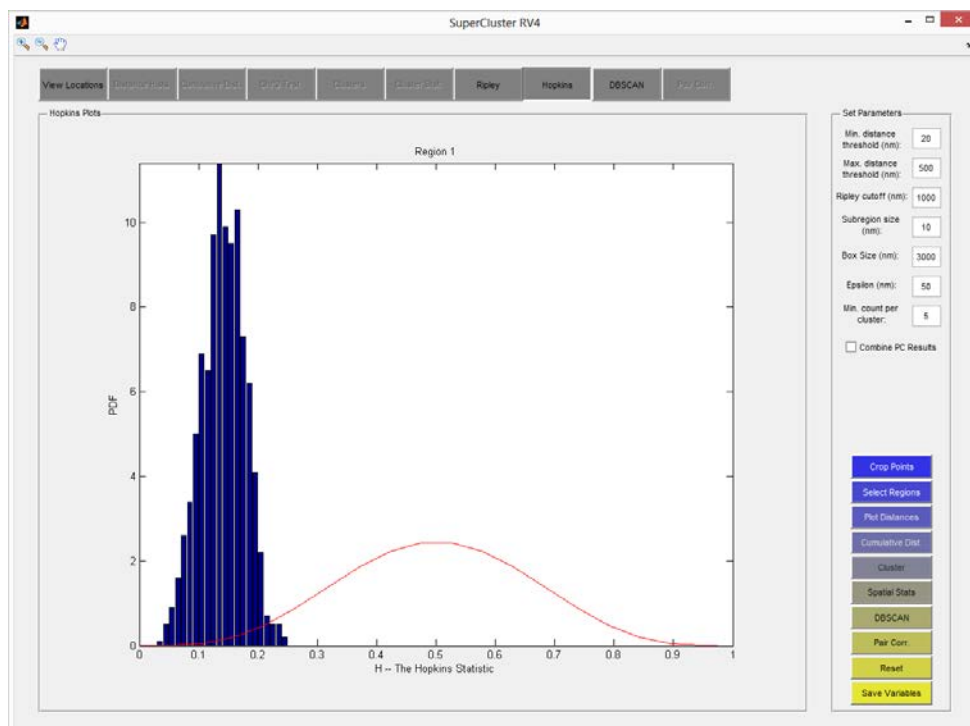


Figura 3.5 – Estatística de Hopkins para amostra uniformemente distribuída no espaço ($H \approx 0$)

A Figura 3.6 representa amostras com tendência a agrupamentos. O gráfico da Figura 3.7 apresenta resultados da Estatística de Hopkins próximos a unidade, conforme valores de H no eixo x.

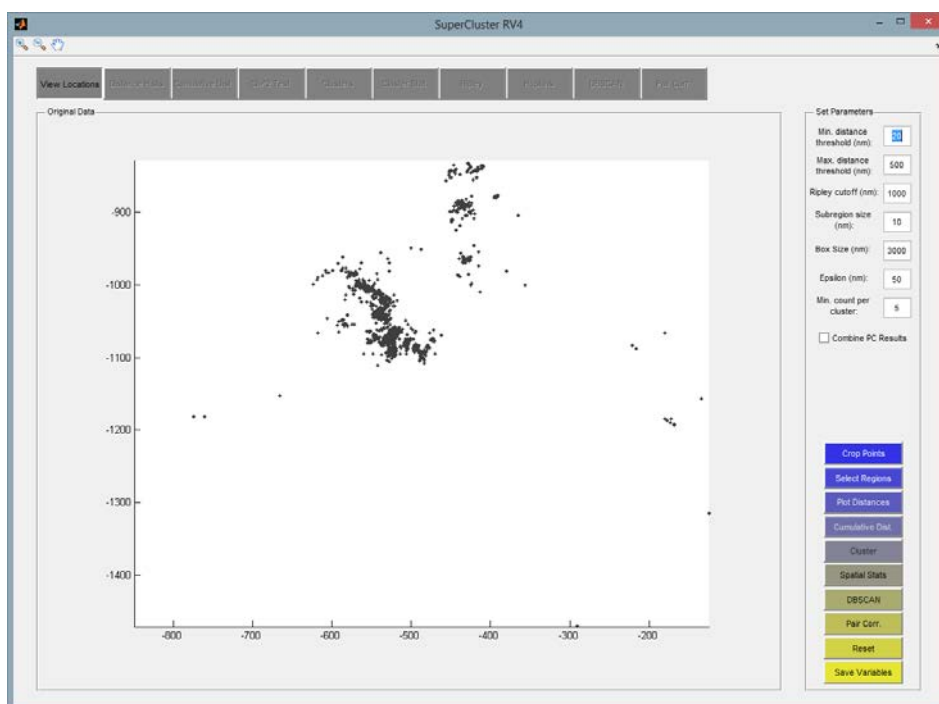


Figura 3.6 – Amostra com dados agrupáveis

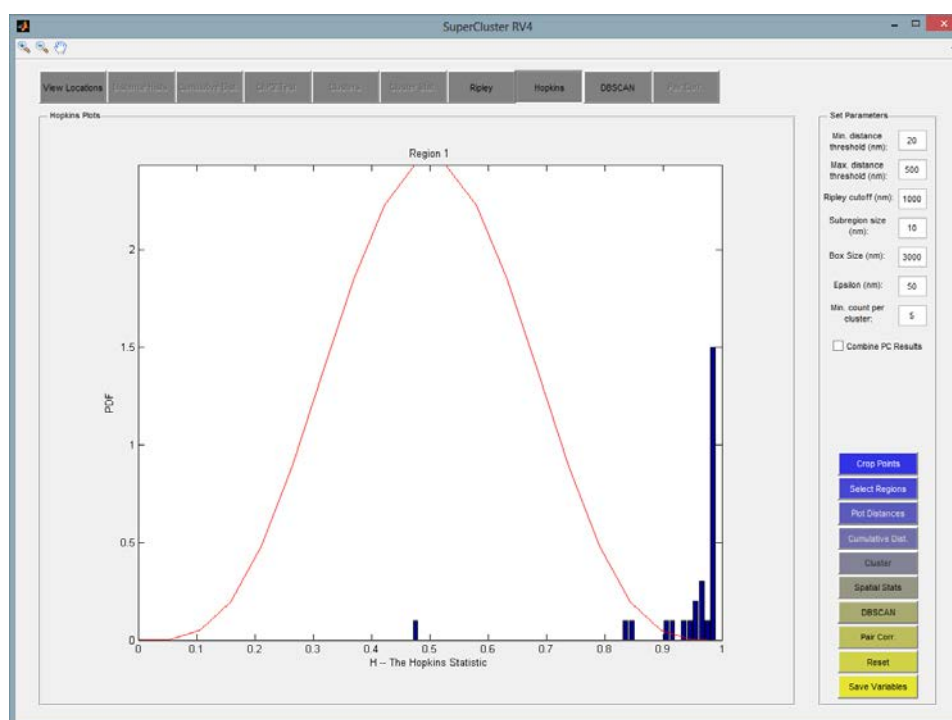


Figura 3.7 – Estatística de Hopkins para amostra com dados agrupáveis ($H \approx 1$)

Foram analisadas a distribuição espacial de amostras de descargas atmosféricas, em uma determinada região geográfica, como por exemplo, a amostra do dia 20/09/2014, representada nas Figuras 3.8 e 3.9 abaixo. Foi possível concluir, através da Estatística de Hopkins, a identificação de uma boa característica de agrupamento nestes tipos de dados ($H \approx 1$).

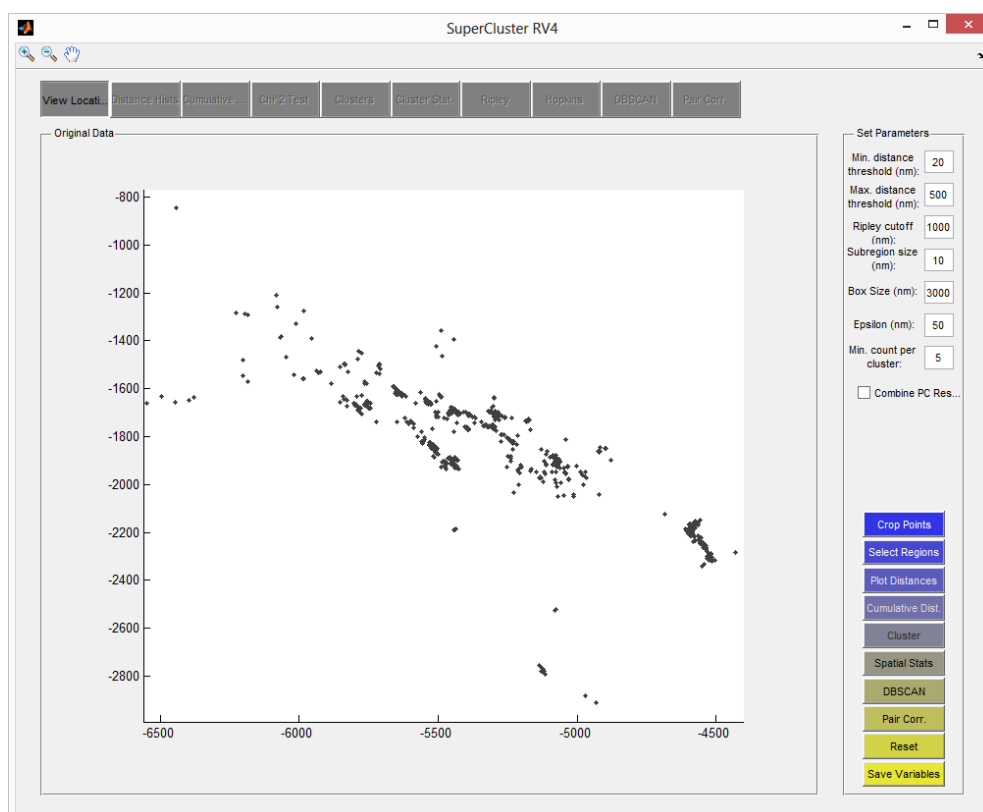


Figura 3.8 – Amostra de descargas atmosféricas dia 20/09/2014

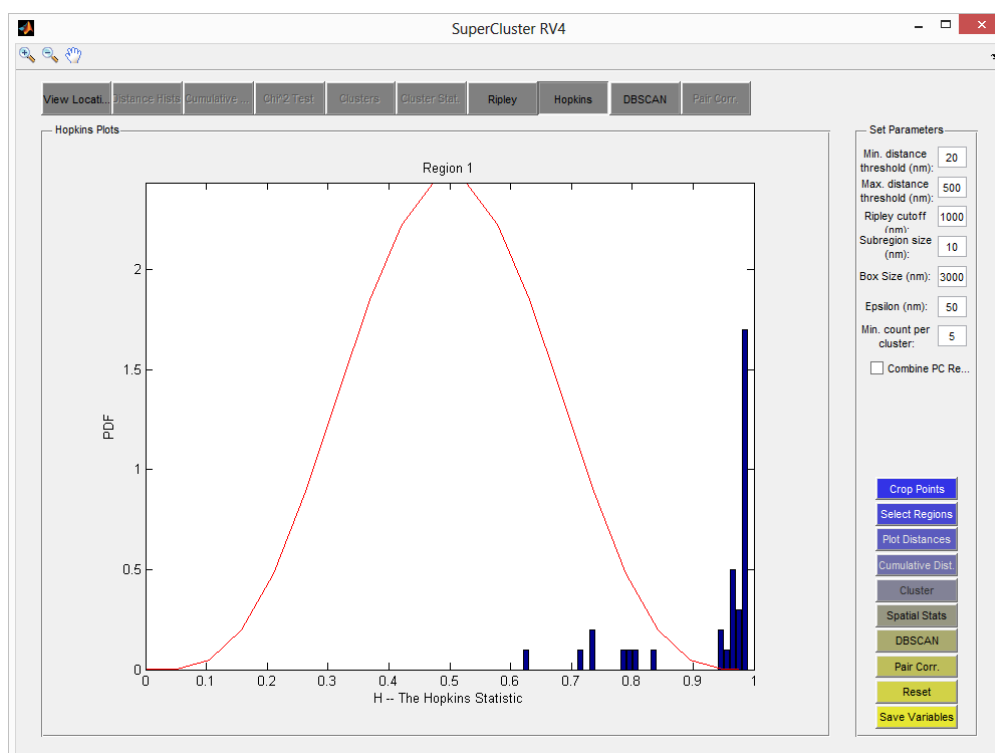


Figura 3.9 – Estatística de Hopkins (descargas atmosféricas 20/9/2014)

3.3.

Dados de Desligamentos de Linhas de Transmissão

O Operador Nacional do Sistema Elétrico (ONS) possui informações consolidadas que relacionam os desligamentos das linhas de transmissão e suas possíveis causas, desde 2002. O universo dessas informações históricas consiste nos desligamentos forçados da Rede de Operação, que é composta da Rede Básica (basicamente instalações com níveis de tensão iguais ou superiores a 230kV), e da Rede Complementar do Sistema Elétrico Brasileiro. De acordo com o Submódulo 2.4 dos Procedimentos de Rede do ONS [28], são estabelecidos limites máximos para taxa de desligamento forçado por descargas atmosféricas, que é o número de desligamentos por descarga atmosférica/100 km/ano, conforme a seguir:

- Circuitos com tensão igual a 230 kV: 2 desligamentos/100 km/ano
- Circuitos com tensão maior ou igual a 345 kV: 1 desligamento/100 km/ano

A partir dessas informações não estruturadas, foi elaborado um banco de dados correlacionando os desligamentos das linhas de transmissão e suas possíveis causas, contendo o nome da linha de transmissão, data e hora do evento, comprimento da linha, nível de tensão e as causas dos desligamentos, segundo códigos definidos pelo ONS, descritos na Tabela 3.2. Os desligamentos em linhas de transmissão são classificados através de 132 causas, e devem ser relacionados, pelas suas áreas competentes, a uma ocorrência no sistema elétrico, após análise da mesma. Esse banco de dados contém 29262 registros sobre desligamentos de linhas de transmissão ocorridos entre os anos de 2002 a 2015, em todo território brasileiro. A elaboração desse banco de dados foi essencial para correlacioná-lo aos dados de descargas atmosféricas, obtidas através das redes de detecção.

Tabela 3.2 – Códigos das causas dos desligamentos das linhas de transmissão

Código	Causa	Código	Causa
A2	Água / Inundação	KJ	Baixo Fluxo de Água de Vedação
A3	Fogo	KK	Baixo Vácuo
AP	Poluição / Contaminação Ambiental	KL	Ferroressonância - Rede
AQ	Queimada / Fogo sob a Linha	KM	Mufla - Falha/Defeito

AS	Depósito Salino	KN	Núcleo - Falha/Defeito
AT	Terremoto	KO	Enrolamento - Falha/Defeito
AU	Umidade	KP	Furo na tubulação
AV	Vibração	KQ	Baixo Fluxo de Óleo
AX	Outras - Meio Ambiente	KR	Baixo Fluxo de Água
DI	Indeterminada	KS	Saturação (TC ou Núcleo de Trafo)
DS	Sem informação do agente	KT	Sobreaquecimento
DV	Vibração	KW	Nível Baixo de Água de Resfriamento
DX	Outras causas	KX	Outras - Equipamentos e Acessórios
F1	Curto-Circuito - Secundário AC	ND	Descarga Atmosférica
F2	Circuito Aberto - Secundário AC	NT	Chuva/Temporal
F3	Oscilação ou Transitório - Secundário AC	NV	Vento Forte
F4	Fusível Queimado (IGN) - Secundário AC	NX	Outros Fenômenos Naturais
F5	Curto-Circuito - Secundário DC	P1	Relé de Proteção - Falha/Defeito
F6	Circuito Aberto - Secundário DC	P2	Instrumento de Medição - Falha/Defeito
F7	Surtos ou Transitório - Secundário DC	P3	Relé Auxiliar AC - Falha/Defeito
F8	Fusível Queimado (IGN) - Secundário DC	P4	Relé Auxiliar DC - Falha/Defeito
F9	Diodo Danificado - Circuito DC	P5	Relé de Proteção Descalibrado
FA	Sobretensão Dinâmica - Circuito DC	P6	Teleproteção - Ruído/Falha/Defeito
FB	Subtensão - Circuito AC/DC	P7	Harmônicas c/ Ruído Suportável na LT
FC	Mistura AC com DC	P8	Ferroressonância no Secundário
FD	Falta de Tensão - Secundário AC/DC	P9	Ruído na Fiação Secundária
FE	Baixa isolamento fiação AC/DC	PA	Falha de Proteção de Outros Componentes
FV	Falta de tensão - AC/DC - Serv. Aux.	PB	Esquema Inadequado de Proteção
FX	Outras - Fiação AC / DC	PC	Atuação Direta Prot. Outros Componentes
FY	Fiação AC/DC - Outra Empresa	PD	Detetor Temperatura - Falha/Defeito
HC	Erro de Operação - Centro de Operação - Agente	PE	Cartão Eletrônico - Falha/Defeito
HD	Erro de operação - Centro de Operação - ONS	PF	Falha de Automatismo
HG	Acidental - Serviços/Testes	PG	Falha no Software PLC
HH	Erro de operação - instalação	PI	Sem Dados - Proteção, Medição, Controle
HJ	Fechamento Fora de Sincronismo	PM	Dispos./Prot. Mecânica - Falha/Defeito
HK	Erro de Relação	PO	Atuação Prot. Outra Empresa Barra Comum
HL	Erro de Fiação DC - Execução	PS	TC Auxiliar - Saturação
HM	Erro de Fiação DC - Projeto	PT	Concepção Inadequada de Teleproteção
HN	Erro de Fiação AC - Execução	PU	Relé Auxiliar do Carrier - Falha/Defeito
HO	Erro de Fiação AC - Projeto	PW	Concepção Inadequada de Relé
HP	Direcionalidade Invertida - Plug	PX	Outras - Proteção, Medição, Controle
HQ	Erro de Ajuste - Cálculo	T1	Condições Anormais de Operação
HR	Erro de Ajuste - Execução	T2	Manobra Operacional Automática (Forçada)
HS	Erro de Configuração de Lógica - Concepção	T5	Ocorr. em Outro Componente da Empresa
HT	Erro de Configuração de Lógica - Implementação	T6	Atuação de Esquema Especial
HX	Outras - Humanas Própria Empresa	TA	Harmônicos em Excesso

HY	Humanas - Outra Empresa	TC	Desligamento por Configuração
K1	Defeito	TI	Sem Dados - Sistema Elétrico
K2	Falha	TO	Deslig. por configuração-outra Empresa
K3	Queima	TV	Falta de Tensão AC/DC – com ou sem Deslig.
K4	Explosão	TX	Outras - Sistema Elétrico
K5	Queda	U1	Perturbação em Outra Empresa Interligada
K6	Rompimento/Desconexão	U2	Perturbação em Consumidor

As informações contendo a correlação entre os registros de desligamento de linhas de transmissão e os dados de descargas atmosféricas, obtidas através das redes de detecção, são exploradas no Capítulo 4 desse estudo, para a definição do espaço amostral em que ocorreram desligamentos de linhas de transmissão causados por descargas atmosféricas, considerando o universo temporal e posicional dos raios.

4

Modelo de Determinação em Tempo Real do Risco de Desligamentos em Linhas de Transmissão por Raios – RDLR

O modelo proposto nesse trabalho para a determinação do risco de desligamentos em linhas de transmissão por raios, chamado por RDLR, é baseado nas ocorrências de descargas atmosféricas que podem afetar as linhas de transmissão, fornecendo informações sobre o risco que apoiem as decisões na operação dos sistemas elétricos para mitigar o impacto indesejável das interrupções no fornecimento de energia elétrica.

O modelo RDLR foi estruturado conforme dois aspectos, sendo o primeiro o contínuo aperfeiçoamento dos resultados na determinação do risco de desligamento, através do processamento de novos registros de desligamento de linhas de transmissão por raios. O segundo aspecto se refere a um modelo dinâmico, baseado nos dados de descargas atmosféricas em tempo real, para a determinação instantânea risco de desligamento de linha de transmissão. Assim, sobre esses dois aspectos, o modelo RDLR foi definido em etapas, que abrangem o pré-processamento dos dados, onde é determinado o espaço amostral de descargas atmosféricas, o agrupamento das mesmas com a definição dos grupos relevantes para processamento e a seleção das variáveis de entrada que representam efetivamente os agrupamentos. Após o pré-processamento das informações, estas são processadas através de uma Rede Neural *Multilayer Perceptron* - MLP, para classificação de padrões de descargas atmosféricas que podem ou não causar desligamentos de linhas de transmissão. Por fim, a partir da classificação obtida da rede MLP, ocorre no módulo decisório a determinação do risco de desligamento de linhas de transmissão por raios.

A Figura 4.1 ilustra o modelo conceitual utilizado para o aprimoramento da determinação do risco.

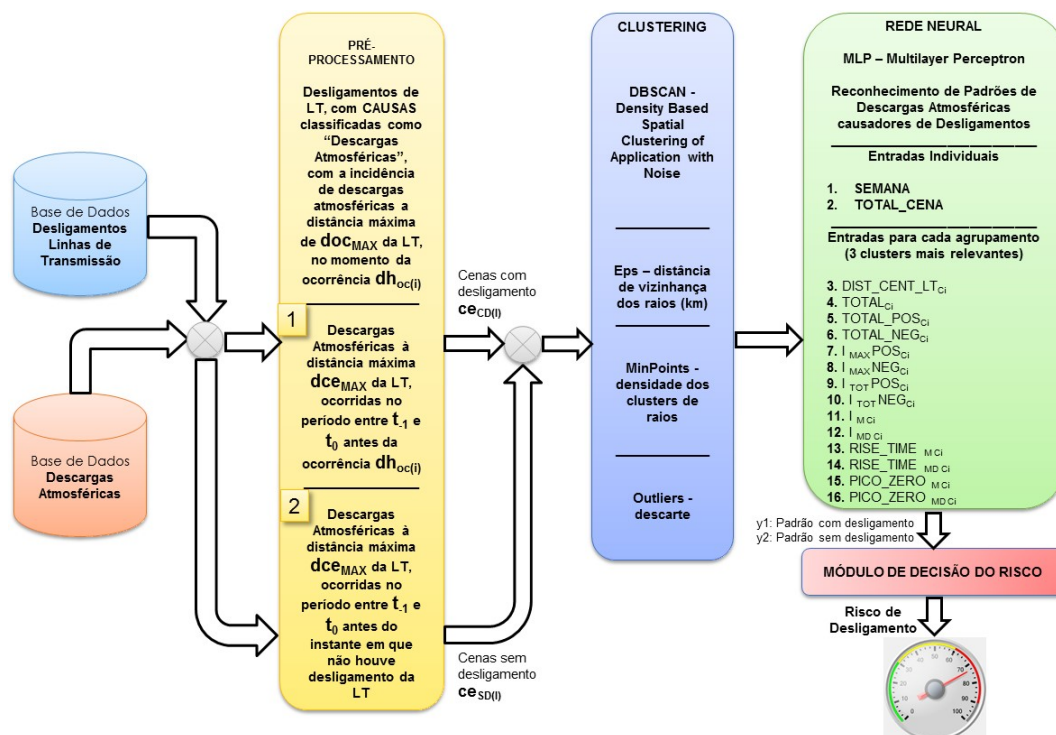


Figura 4.1 – Modelo RDLR Conceitual – Otimização dos Parâmetros

4.1. Pré-processamento dos Dados

4.1.1. Definição do Espaço Amostral

Na etapa do pré-processamento dos dados, inicialmente é realizada a validação da ocorrência de desligamento da linha de transmissão devido à incidência de descargas atmosféricas, a uma distância máxima (d_{oc_MAX}) da linha de transmissão, no momento da ocorrência (dh_{oc}). Caso afirmativo, são extraídas amostras de descargas atmosféricas relacionadas à linha de transmissão, considerando os instantes com desligamentos e instantes sem a ocorrência de desligamentos, definidas neste trabalho como cenas de descargas atmosféricas. A definição de uma cena de raios que antecede o instante atual (t_0) é determinante para que o modelo RDLR forneça informações antecipadas sobre riscos de desligamentos de uma linha de transmissão em tempo real. Dessa forma, o modelo RDLR analisa eventos de descargas atmosféricas que ocorreram na janela temporal ($t_1 - t_0$), no instante t_0 antes da ocorrência dos desligamentos (dh_{oc}), onde t_1 e t_0 correspondem, respectivamente, ao início e fim da janela de

análise, a pelo menos a distância $d_{ce_{MAX}}$ da linha de transmissão em questão. A definição do espaço amostral estudado nesse modelo permite observar cenários característicos de raios que são responsáveis pelo desligamento de linhas de transmissão (ce_{CD}). Essas cenas são selecionadas através da observação da incidência de descargas atmosféricas a distância máxima ($do_{C_{MAX}}$) da linha de transmissão. Além desses cenários, o modelo também analisa amostras de descargas atmosféricas não associadas a desligamentos de linhas de transmissão (ce_{SD}).

As cenas definidas, normalmente são caracterizadas por possuírem grande volume de dados, sendo necessário o agrupamento desse conjunto de dados.

4.1.2.

Agrupamento dos Dados de Descargas Atmosféricas das Cenas

Após a definição das cenas com os dados de descargas atmosféricas, relacionado à uma linha de transmissão, a amostra possui um grande conjunto de dados, que deverá ser agrupado. Nesse estudo, o algoritmo DBSCAN (*Density Based Spatial Clustering of Application with Noise* (Agrupamento Espacial Baseada em Densidade de Aplicações com Ruído) foi utilizado como o método de agrupamento dos dados, por ser capaz de identificar grupos de formatos e tamanhos distintos, bem como separação espacial eficiente de ruídos (*outliers*).

A maioria dos métodos particionais e hierárquicos agrupam objetos baseando-se na distância entre eles. Por isso, tais métodos podem encontrar dificuldades para descobrir grupos de formas distintas. Nos métodos de agrupamento baseados em densidade, grupos são definidos como regiões densas, separadas por regiões menos densas que representam ruídos. As regiões densas podem ter diferentes formas e os pontos dentro de uma região podem também estar distribuídos distintamente e, por isso, estes métodos são indicados para filtrar ruídos e descobrir grupos com formas distintas, tais como elíptica, cilíndrica, espiralada, etc. (HAN e KAMBER, 2001).

O DBSCAN encontra agrupamentos baseando-se na vizinhança dos objetos, onde a densidade associada a um ponto é obtida por meio da contagem do número de pontos vizinhos em uma determinada região ao redor desse ponto [ERTÖZ, STEINBACH; KUMAR, 2003]. Esse algoritmo possui a capacidade de encontrar agrupamentos considerando as propriedades dos dados, pois não requer que seja informado antecipadamente o número de grupos, permitindo a

formação de agrupamentos de diferentes formas, sendo bastante apropriado para criar formações geográficas de tempestades a partir dos dados posicionais das descargas atmosféricas. Em contrapartida, são necessários outros dois parâmetros de entrada para o algoritmo, tal como raio de vizinhança de um ponto (Eps) e número mínimo de pontos (MinPts). Outras características importantes do algoritmo são a capacidade de identificar ruídos (*outliers*) e a possibilidade de poder trabalhar com diversas funções de distância [ANKERST et al, 1999; METZ, 2006; ESTER et al, 1996].

A Tabela 4.1 apresenta os parâmetros utilizados no modelo, envolvendo o algoritmo DBSCAN e os critérios de seleção das cenas nesse estudo.

Tabela 4.1 – Parâmetros do Modelo Conceitual

#	Parâmetro	Descrição
1	dce_{MAX}	Distância máxima em metros entre a linha de transmissão e as descargas atmosféricas observadas na Cena(i)*
2	t_0	Tempo máximo em minutos que antecede o momento da ocorrência de desligamento, suficiente para a realização de ações operativas (reprogramações antecipadas) para reduzir o impacto do desligamento da LT. Fim da janela temporal da Cena(i).
3	t_1	Início da janela temporal de descargas atmosféricas da Cena(i), representada por $(t_1 - t_0)$.
4	doc_{MAX}	Distância máxima em metros entre as descargas atmosféricas e a linha de transmissão utilizada para a seleção das cenas
5	$dh_{oc(i)}$	Data e horas das cenas que ocorreram desligamento (i) das linhas de transmissão
6	$ce_{CD(i)}$	Cenas(i) que ocorreram desligamentos que atenderam ao critério de seleção
7	$ce_{SD(i)}$	Cenas(i) que não ocorreram desligamentos, onde foram observadas a incidência de descargas atmosféricas, aleatoriamente selecionadas dentro da amplitude de desligamentos ($dh_{oc(i)}$).
8	Eps	Distância de vizinhança dos raios para o processo de clusterização
9	MinPoints	Densidade dos clusters de raios

As linhas de transmissão possuem dispositivos de proteção que garantem seu funcionamento até um determinado nível de exposição à incidência de raios, e quase sempre os desligamentos associados a essa natureza dependem de uma incidência massiva de raios sobre as linhas, sendo o estudo destes agrupamentos fundamental para a determinação do risco de desligamentos. Conforme descrito no Capítulo 2, dentre os parâmetros para estudos de desempenho de sistemas elétricos, a densidade de descargas para o solo é um dos mais importantes.

4.1.3.

Seleção dos Agrupamentos de Descargas Atmosféricas Relevantes

Seguindo o agrupamento dos dados, o Modelo RDLR define aqueles mais relevantes, de forma a garantir melhores resultados e maior desempenho da rede. Dessa forma, o critério que considera a razão entre a quantidade de raios e distância do centroide do grupo à linha de transmissão é o método proposto nesse estudo para selecionar os grupos que possuem maior representatividade no conjunto de dados de uma amostra. O modelo RDLR permite considerar um número arbitrário de agrupamentos para serem processados através da Rede Neural MLP.

Cada descarga atmosférica que compõe os agrupamentos possui informações individuais que caracterizam efetivamente os seus grupos. As variáveis de entrada consideradas são: a distância do centroide do grupo à linha de transmissão ($DIST_CENT_LT_{Ci}$), o total de raios do grupo ($TOTAL_{Ci}$), número de descargas com polaridade positiva ($TOTAL_POS_{Ci}$), número de descargas com polaridade negativa ($TOTAL_NEG_{Ci}$), corrente máxima positiva ($I_MAX_POS_{Ci}$) e negativa do grupo ($I_MAX_NEG_{Ci}$), corrente total positiva ($I_TOT_POS_{Ci}$) e negativa ($I_TOT_NEG_{Ci}$), corrente média (I_M_{Ci}) e mediana (I_MD_{Ci}), tempo de ascensão médio ($RISE_TIME_M_{Ci}$) e mediano ($RISE_TIME_MD_{Ci}$), tempo de pico a zero médio ($PICO_ZERO_M_{Ci}$) e mediano ($PICO_ZERO_MD_{Ci}$) do grupo. Além dessas variáveis de entrada, são consideradas também a sazonalidade ($SEMANA$) e a quantidade de raios ($TOTAL_CENA$) do evento, totalizando 16 variáveis de entrada em cada agrupamento.

4.1.4.

Seleção das Variáveis de Entrada por Análise de Componentes Principais

A Análise de Componentes Principais – PCA (*Principal Component Analysis*) foi primeiro introduzida por Pearson em 1901 e desenvolvida de forma independente por Hotelling em 1933 [30].

A PCA consiste essencialmente em uma transformação de coordenadas de um conjunto de variáveis originais correlacionadas em um conjunto de variáveis sintéticas descorrelacionadas, as chamadas Componentes Principais (CP). Cada componente é uma combinação linear das variáveis originais. Todas as

Componentes Principais são ortogonais entre si, portanto não há informações redundantes.

No primeiro passo, os dados são organizados em uma matriz $X_{(n \times m)}$, onde n representa o número de observações e m o número de variáveis independentes.

O segundo passo, a centralização dos dados em torno da média, tem como objetivo prevenir que os pontos mais distantes do centro dos dados tenham maior influência que os mais próximos, o que é apropriado para dados em diferentes unidades [32]. Essa transformação é feita pela aplicação da seguinte equação:

$$Z_i = \frac{(X_i - \mu)}{\sigma} \quad (3)$$

Onde Z_i são os valores transformados, X_i são as amostras para $i = 1, 2, \dots, n$, μ a média amostral da variável aleatória x_i e σ é o desvio padrão amostral. Desta forma, todas as variáveis aleatórias são distribuídas com média zero e desvio padrão unitário. O terceiro passo consiste no cálculo da matriz de covariância C_z :

$$C_z = Z \cdot Z^T \quad (4)$$

Onde, Z^T é a transposta de Z . Os elementos da diagonal principal desta matriz se referem às variâncias (λ) das colunas (variáveis independentes). Já os elementos fora da diagonal principal representam a covariância entre as variáveis. Quando a covariância é nula significa que as variáveis aleatórias são descorrelacionadas. A matriz de covariância é real e simétrica, o que significa ser possível encontrar um conjunto de autovalores e correspondentes autovetores ortogonais.

O quarto passo consiste então na determinação dos autovalores λ e autovetores v_n correspondentes da matriz C_z . Os autovetores são dispostos de modo decrescente, de acordo com os valores dos autovalores. Encontrados os autovetores v_n , estes formarão as colunas de uma matriz P :

$$P = (v_1, v_2, \dots, v_n) \quad (5)$$

Em PCA, os autovalores da matriz de covariância correspondem à variância das variáveis transformadas. Assim, se um autovetor possui autovalor grande, significa que este fica em uma direção em que há uma grande variância dos

padrões. A importância desse fato está na classificação, pois é mais fácil distinguir padrões usando uma base em que seus vetores não são correlacionados e que apontam para a direção da maior variância dos dados.

Os elementos dos autovetores, chamados de *loadings* na terminologia PCA, representam os cossenos diretores, ou seja, a contribuição com que cada um dos eixos originais entra na composição dos novos eixos, que são as Componentes Principais. Por sua vez, os autovalores representam a quantidade de variância original descrita pelos respectivos autovetores, denominados por *scores*. Os *scores* são as projeções das amostras na direção das componentes principais. Suas coordenadas são obtidas a partir do produto da matriz de dados pela matriz de autovetores [31].

O último passo é a diagonalização. A matriz P é empregada para mudar a base de C_z obtendo uma matriz diagonal D de autovalores de C_z .

$$D = P^{-1}C_zP \quad (6)$$

A matriz D apresenta elementos iguais aos autovalores na diagonal principal, ou seja, não apresenta covariância e conseqüentemente, não tem nenhuma informação redundante [32].

Na nova base formada pela PCA, um menor número de coordenadas é suficiente para conter grande parte da informação explicada pela distribuição dos dados, ou seja, as primeiras Componentes Principais apresentam grande variância dos dados, enquanto outras têm variância próxima de zero. Assim, quanto maior a porcentagem de variância nos dados puder ser explicada pelas primeiras Componentes Principais, uma redução de dimensionalidade mais eficiente pode ser obtida. A porcentagem de variância total explicada por cada Componente Principal é representada pela razão da variância (autovalor) correspondente e da soma total das variâncias.

Apesar do gráfico dos *scores* analisar apenas duas Componentes Principais ao mesmo tempo, esta parece ser uma alternativa interessante para identificação de possíveis *outliers*, já que se trata das componentes com maior variabilidade explicada dos dados originais.

O elevado grau de complexidade presente nas Redes Neurais Artificiais pode prejudicar seu desempenho, e a presença de variáveis de entrada redundantes, que nada acrescentam ao seu processo de aprendizagem, tornam o treinamento mais difícil e demorado. O método PCA foi aplicado nesse estudo

para selecionar quais variáveis de entrada possuem características mais relevantes para a determinação da saída ou resposta da rede, possibilitando a redução do número de entradas. O método proposto nesse trabalho seleciona as variáveis de entrada através da raiz quadrada da soma dos quadrados dos *loadings* calculados pelo método PCA.

4.2.

O Modelo da Rede Neural MLP

4.2.1.

Variáveis de Entrada

As variáveis de entrada descritas na Tabela 4.2 são valores que representam os agrupamentos de raios, de acordo com os atributos das descargas atmosféricas, descritos no Capítulo 3, item 3.1 desse trabalho. São definidas duas variáveis de entrada que tratam da semana atual e total de raios existentes na cena. As demais variáveis de entrada se referem aos agrupamentos de descargas atmosféricas, sendo compostas por 14 variáveis de entrada para cada grupo.

Tabela 4.2 – Descrição das Variáveis de Entrada

Entradas		Variável	Descrição
Indiv- duais	1	SEMANA	Semana do Evento (1..52)
	2	TOTAL_CENA	Total de Raios da Cena
Cluster (i)	3	DIST_CENT_LT _{Ci}	Distância do Centróide do Cluster (i) à Linha de Transmissão - m
	4	TOTAL _{Ci}	Total de Raios do Cluster (i)
	5	TOTAL_POS _{Ci}	Total de Raios Positivos no Cluster (i)
	6	TOTAL_NEG _{Ci}	Total de Raios Negativos no Cluster (i)
	7	I _{MAX} POS _{Ci}	Corrente Máxima dos Raios Positivos no Cluster (i) - kA
	8	I _{MAX} NEG _{Ci}	Corrente Máxima dos Raios Negativos no Cluster (i) - kA
	9	I _{TOT} POS _{Ci}	Corrente Total dos Raios Positivos no Cluster (i) - kA
	10	I _{TOT} NEG _{Ci}	Corrente Total dos Raios Negativos no Cluster (i) - kA
	11	I _M Ci	Corrente Média dos Raios no Cluster (i) - kA
	12	I _{MD} Ci	Corrente Mediana dos Raios no Cluster (i) - kA
	13	RISE_TIME _M Ci	Tempo de Ascensão Médio no Cluster (i) - ms
	14	RISE_TIME _{MD} Ci	Tempo de Ascensão Mediano no Cluster (i) - ms
	15	PICO_ZERO _M Ci	Tempo de Pico a Zero Médio no Cluster (i) - ms
	16	PICO_ZERO _{MD} Ci	Tempo de Pico a Zero Mediano no Cluster (i) - ms

A quantidade de neurônios na camada de entrada da Rede Neural MLP é definida, conforme a equação a seguir:

$$In = 2 + (14 \times Nc) \quad (7)$$

Onde

In = número de neurônios na camada de entrada

Nc = número de agrupamentos selecionados

O Modelo RDLR foi estruturado de forma a definir diferentes quantidades de elementos na camada de entrada, possibilitando o treinamento, validação e teste da Rede Neural MLP considerando todas as variáveis de entrada ou possibilitando uma redução das mesmas, selecionadas através do método PCA, descrito anteriormente.

4.2.2. A Camada Intermediária

O Modelo RDLR define 8 neurônios presentes na camada intermediária, que resultam em um menor erro médio sobre o conjunto de teste, definido após a execução de 50 instâncias no treinamento da Rede Neural MLP. Esse número de

elementos na camada intermediária garante o melhor ajuste da modelagem dos dados e capacidade de aproximação da Rede Neural MLP.

O número de neurônios na camada intermediária é uma importante questão no projeto da arquitetura de uma Rede Neural MLP. Uma grande quantidade de elementos pode levar a rede a uma diminuição da capacidade de generalização (*overfitting*), impedindo a extração das características gerais. Por outro lado, um número muito pequeno de neurônios pode tornar a estrutura insuficiente para manipular as entradas, dificultando sobremaneira a convergência dos algoritmos de treinamento (*underfitting*).

4.2.3. As Variáveis de Saída

A Rede Neural MLP do Modelo RDLR é estruturada através de duas classes, definidas como variáveis de saída da rede. Uma classe representa padrões de agrupamentos de descargas atmosféricas responsáveis por desligamento de linhas de transmissão (y_1) e a outra classe representa os agrupamentos de descargas atmosféricas não associados a desligamento de linhas de transmissão (y_2), conforme Tabela 4.3.

Tabela 4.3 – Descrição das Variáveis de Saída

Saídas	Variável	Descrição
1	y_1	Padrão com desligamento de LT por raios (0..1)
2	y_2	Padrão sem desligamento de LT por raios (0..1)

São executadas várias interações de treinamento, validação e teste, sendo que para cada uma delas, a rede neural calcula o valor das variáveis que representam as classes, referente a cada uma das amostras.

O valor de cada uma das variáveis de saída representa a capacidade classificatória da rede, permitindo entender padrões de descargas atmosféricas que de fato possam causar perturbações nas linhas de transmissão.

A Figura 4.2 ilustra a arquitetura da Rede Neural MLP do Modelo RDLR.

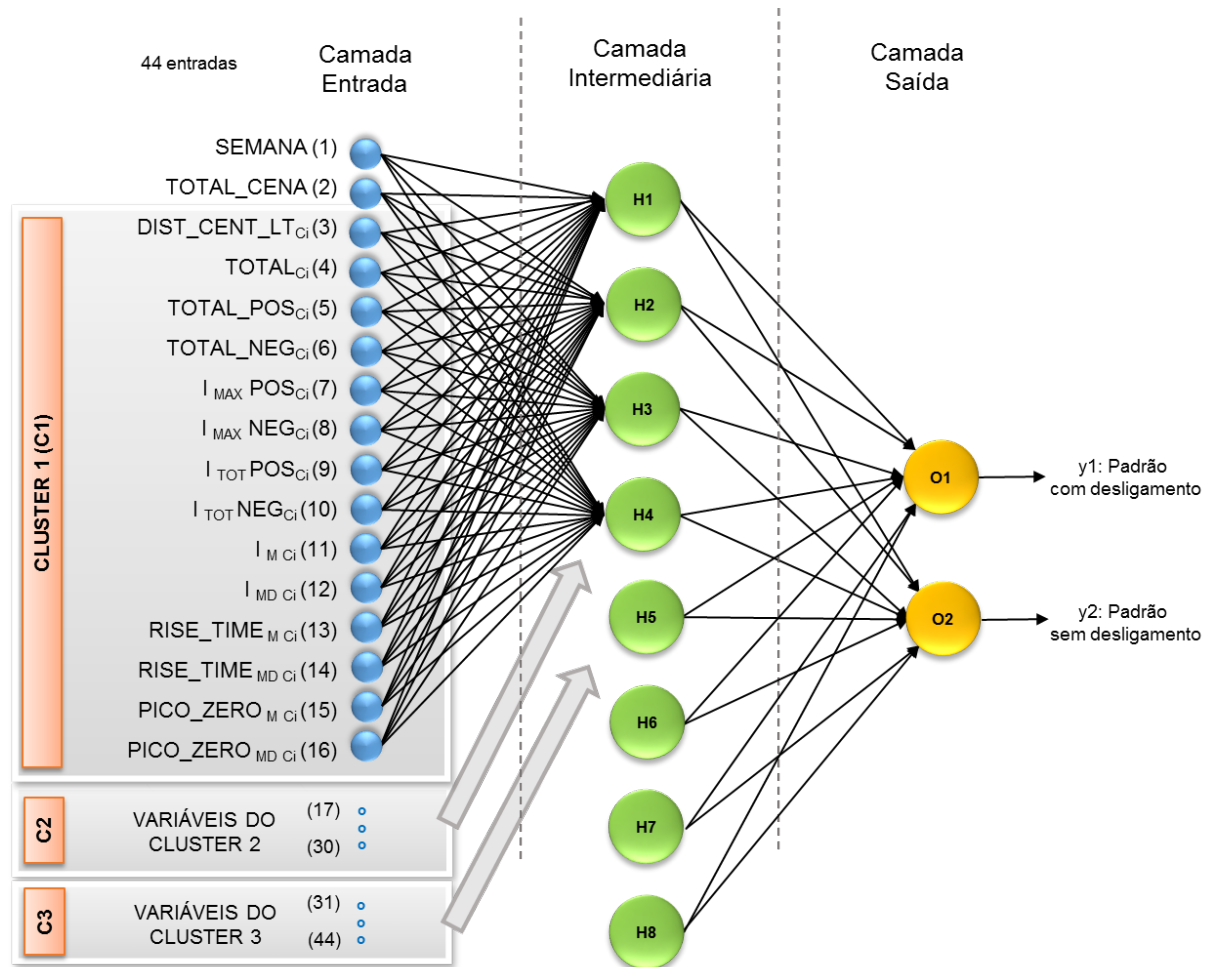


Figura 4.2 - A Rede Neural MLP do Modelo RDLR

4.3.

Definição da Determinação do Risco em Tempo Real do Desligamento em Linhas de Transmissão por Descargas Atmosféricas

A partir dos valores das variáveis de saída da rede neural MLP, o Modelo RDLR definiu um módulo decisório para determinar o risco do desligamento de linhas de transmissão por descargas atmosféricas. Assim, a Rede Neural MLP classificará o padrão de descargas atmosféricas de uma determinada cena, e a classe que possui maior valor absoluto é escolhida, e a partir dela o cálculo do risco é produzido.

O risco RDLR é definido, conforme condição abaixo:

Se $y_1 > y_2$, então

$$RDLR(\%) = y_1 \times 100\% \quad (8)$$

senão

$$RDLR(\%) = (1 - y_2) \times 100\% \quad (9)$$

onde,

RDLR - risco do desligamento da linha de transmissão

y_1 – variável de saída do padrão de desligamento da LT (classe 1)

y_2 – variável de saída do padrão não associado a desligamento da LT (classe 2)

4.4. Modelo RDLR Dinâmico (Tempo Real)

O Modelo RDLR dinâmico pode ser integrado a sistemas computacionais de apoio à tomada de decisão, seguindo os conceitos da Consciência Situacional, pois possibilita a percepção dos elementos do ambiente, a compreensão do seu significado e a projeção de seu estado no futuro próximo [34].

Com a utilização do modelo na forma dinâmica, a determinação do risco de desligamento por descargas atmosféricas é informada em tempo real, para qualquer linha de transmissão representada geograficamente, a partir dos dados de descargas atmosféricas coletados através das redes de detecção em tempo real, com a geração das cenas de descargas atmosféricas a uma distância máxima d_{ceMAX} da linha de transmissão (LT), ocorridas entre uma janela móvel de tempo, compreendida entre os instantes t_{-1} e t_0 (atual). A determinação do risco é fornecida para um instante dh_{OC} , posterior ao instante t_0 atual, para permitir aos operadores realizarem reprogramações do sistema elétrico interligado, através do remanejamento do fluxo de energia para outras linhas que não se encontrem sob risco, reduzindo ou até mesmo anulando o impacto no caso de um desligamento por descargas atmosféricas.

Após a definição do espaço amostral são gerados os agrupamentos dos dados pelo algoritmo DBSCAN, seguindo a seleção dos grupos mais

representativos da cena. Os agrupamentos são processados na Rede Neural MLP, que classificará o padrão de descargas típico de desligamento de linhas de transmissão ou um padrão típico de não desligamento de linhas de transmissão.

Com base na classificação obtida da Rede Neural MLP, o módulo de decisão RDLR determinará o risco de desligamento da linha de transmissão, fornecendo informações que permitam ações operativas para a manutenção da segurança do sistema elétrico. A Figura 4.3 representa os módulos do Modelo RDLR dinâmico.

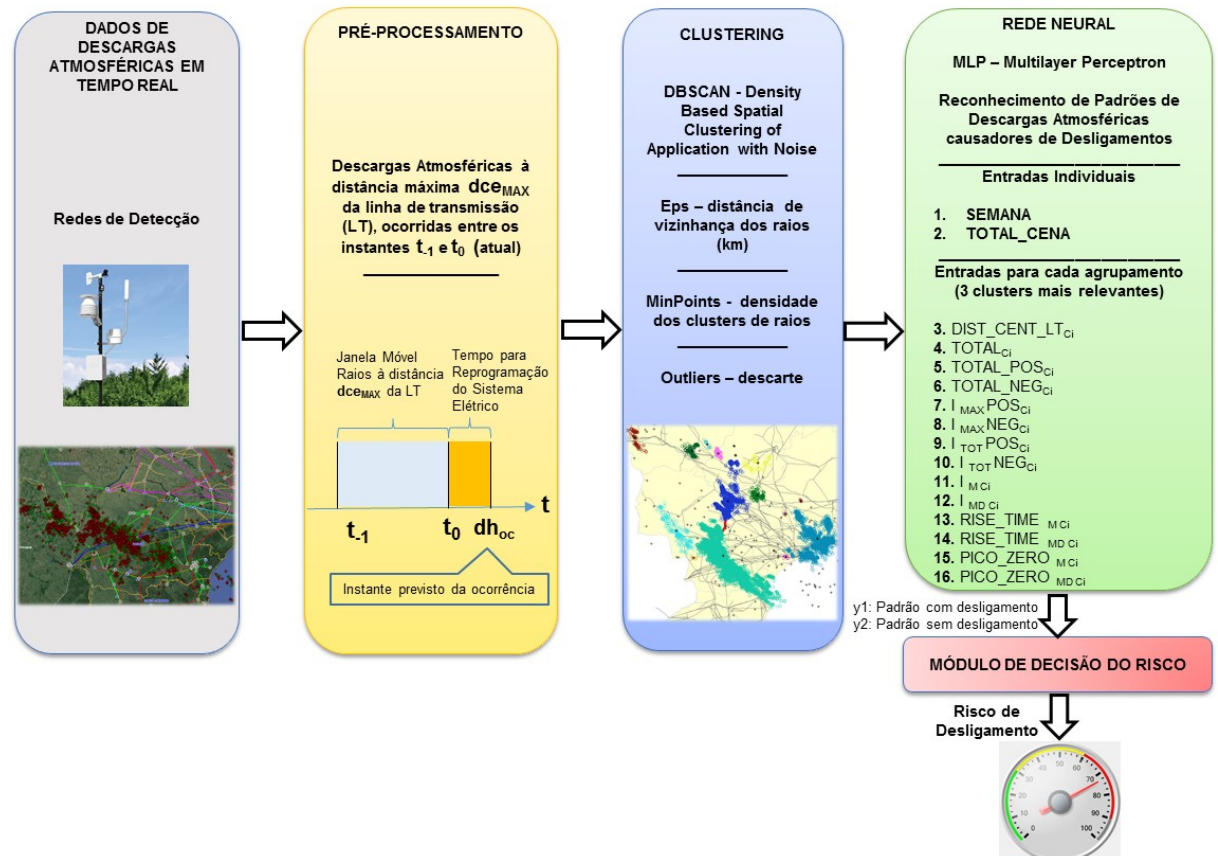


Figura 4.3 – Modelo RDLR Dinâmico (Tempo Real)

5 Estudo de Caso

5.1. Análise dos Dados de Desligamento de Linhas de Transmissão

Nesse estudo foi realizada uma análise dos dados com 29262 registros sobre desligamentos de linhas de transmissão, relacionados ao banco de dados de descargas atmosféricas, obtidos através das redes de detecção, que contém aproximadamente 400 milhões de registros. O banco de dados permitiu extrair informações sobre as causas associadas aos desligamentos das linhas de transmissão que ocorrem mais frequentemente e associá-las a incidência de descargas durante estes desligamentos.

A Tabela 5.1 e a Figura 5.1 apresentam a classificação dos desligamentos de linhas de transmissão e suas causas correlacionadas à quantidade de descargas atmosféricas registradas entre 2002 e 2015. É possível observar um grande número de desligamentos classificados por causa indeterminada. Porém, o monitoramento dos desligamentos forçados depende da correta identificação das causas dos desligamentos pelas transmissoras. Por este motivo, é muito importante que as transmissoras aprimorem seus processos de identificação de perturbações e desligamentos forçados. O objetivo é que seja reduzido o número de desligamentos com causas indeterminadas, permitindo um monitoramento mais eficaz das causas dos desligamentos no sistema de transmissão.

Tabela 5.1 - Dados sobre desligamentos de linhas de transmissão e quantidade de descargas atmosféricas

Causas de Desligamentos	Quantidade (2002 a 2015)	Percentual	Quantidade Acumulada	Percentual Acumulada
Indeterminada	8103	28%	8103	28%
Descarga Atmosférica	5809	20%	13912	48%
Queimada / Fogo sob a Linha	2723	9%	16635	57%
Atuação Direta Prot. Outros Componentes	1506	5%	18141	62%
Chuva/Temporal	1100	4%	19241	66%
Vegetação	940	3%	20181	69%
Acidental - Serviços/Testes	913	3%	21094	72%
Ocorr. em Outro Componente da Empresa	704	2%	21798	74%
Condições Anormais de Operação	531	2%	22329	76%
Teleproteção - Ruído/Falha/Defeito	489	2%	22818	78%
Vento Forte	430	1%	23248	79%
Relé de Proteção - Falha/Defeito	408	1%	23656	81%
Atuação de Esquema Especial	383	1%	24039	82%
Perturbação em Outra Empresa Interligada	305	1%	24344	83%
Erro de Ajuste - Cálculo	289	1%	24633	84%
Rompimento/Desconexão	240	1%	24873	85%
Pássaro Curicaca	210	1%	25083	86%
Perturbação em Concessionária Reg/Mun	197	1%	25280	86%
Poluição / Contaminação Ambiental	181	1%	25461	87%
Falha	172	1%	25633	88%
Defeito	172	1%	25805	88%
Desligamento por Configuração	172	1%	25977	89%
Atuação Prot. Outra Empresa Barra Comum	165	1%	26142	89%
Isolação - Falha/Defeito	156	1%	26298	90%
Animais, Pássaros, Insetos	138	0%	26436	90%
Falha de Proteção de Outros Componentes	122	0%	26558	91%
Baixa Pressão de Ar ou SF6	108	0%	26666	91%
Queda	107	0%	26773	91%
Explosão	104	0%	26877	92%

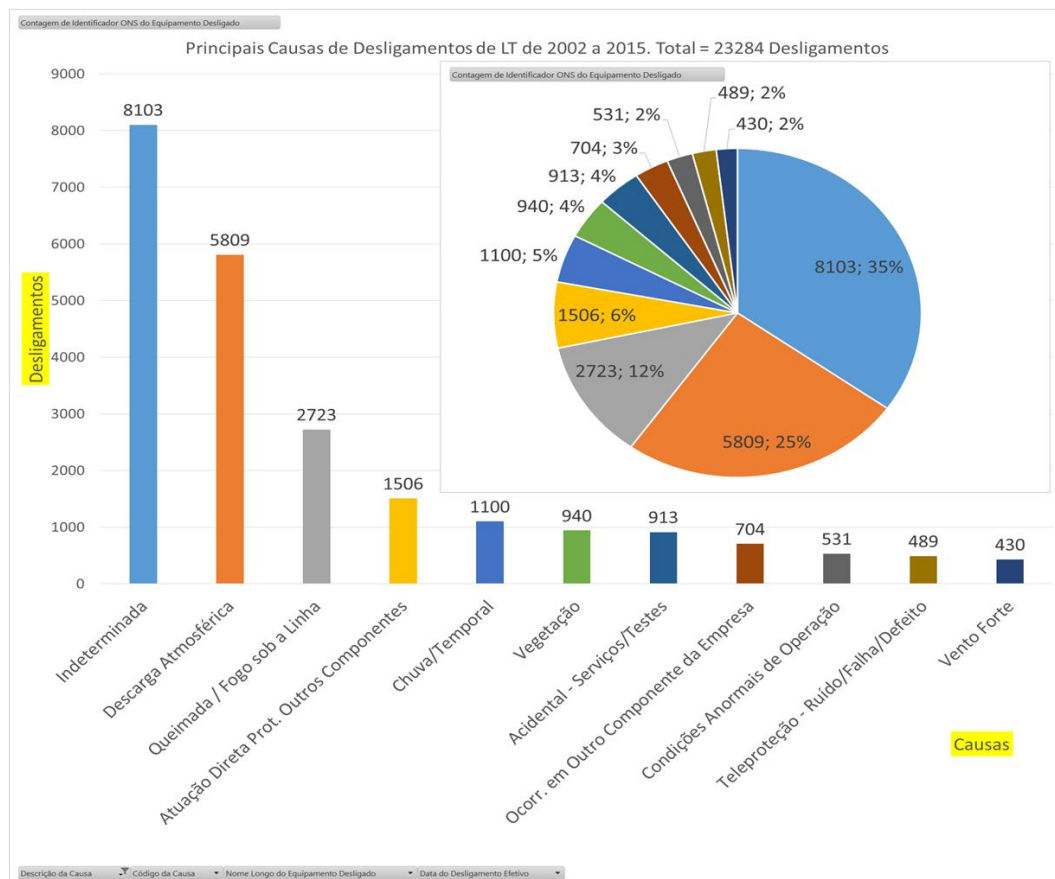


Figura 5.1 - Principais causas de desligamentos de linhas de transmissão – 2002 a 2015

5.2. Pré-processamento dos dados de desligamento de linhas de transmissão e descargas atmosféricas

Após a seleção inicial dos dados de desligamentos causados por descargas atmosféricas, foi necessário realizar um pré-processamento desses dados, para a seleção da linha de transmissão a ser estudada nesse trabalho e, em seguida, a definição do conjunto amostral de descargas atmosféricas, para seu tratamento através do modelo de determinação do risco de desligamento de linhas de transmissão.

5.2.1. Critério de Seleção das Linhas de Transmissão para Estudo

A seleção adequada de uma linha de transmissão para análise do padrão de ocorrência de descargas atmosféricas que causaram o desligamento desse

equipamento é importante para garantir uma consistência do modelo proposto nesse trabalho. Para tal, foram adotados critérios para conduzir a uma boa indicação da linha de transmissão para o estudo. Dessa forma, foram analisados apenas os desligamentos de linhas de transmissão classificados como “descargas atmosféricas”, conforme causa com código ND da Tabela 2 do Capítulo 3. Essa premissa visa garantir a presença de raios nas proximidades da linha de transmissão em questão, observada no banco de dados oriundos das redes de detecção de raios.

A partir dessa conceituação e com base na data e hora dos registros de desligamentos, foram apuradas as quantidades de descargas atmosféricas para cada uma dessas ocorrências, a pelo menos 1 km (doc_{MAX}) das linhas de transmissão, 1 hora antes e 1 hora após o momento do desligamento (dh_{oc}). A Tabela 5.2 expressa o número de descargas atmosféricas através do cruzamento dos bancos de dados de ocorrência de desligamentos (causa de desligamentos) e de incidência de descargas atmosféricas (coletados pelas redes de detecção), agregado por nível de tensão das linhas de transmissão, conforme critério definido.

Tabela 5.2 - Número de descargas atmosféricas por nível de tensão, no critério definido

Indeterminada	8103
Descarga Atmosférica	5809
765	82
600	6
525	245
500	1389
440	284
345	707
230	3023
138	73
Chuva/Temporal	1100
Vegetação	940
Vento Forte	430
Queda	107
Água / Inundação	21
Causa a ser Determinada	18
Total Geral	16528

É observado um grande número de ocorrências associadas ao setor de 500kV, contendo um total 1389 descargas a menos de 1 km das linhas de

transmissão, no intervalo de 1 hora antes e 1 hora após os desligamentos, conforme apresentado na Tabela 5.3.

Tabela 5.3 - Quantidade de desligamentos no setor 500 kV

500			1389
ND			1389
⊕ LT 500 kV GURUPI	/ MIRACEMA	C 1 TO	92
⊕ LT 500 kV S.DA MESA	/ SAMAMBAIA	C 3 GO/DF	68
⊕ LT 500 kV COLINAS	/ MIRACEMA	C 1 TO	59
⊕ LT 500 kV S.DA MESA	/ GURUPI	C 1 GO/TO	48
⊕ LT 500 kV JAGUARA-SE	/ SAO SIMAO-SE	C 1 MG	47
⊕ LT 500 kV RIO DAS EGUAS	/ B.J.LAPA II	C 1 BA	46
⊕ LT 500 kV IMPERATRIZ	/ COLINAS	C 1 MA/TO	46
⊕ LT 500 kV S.DA MESA	/ GURUPI	C 2 GO/TO	43
⊕ LT 500 kV GURUPI	/ MIRACEMA	C 3 TO	40
⊕ LT 500 kV S.DA MESA	/ SAMAMBAIA	C 2 GO/DF	36
⊕ LT 500 kV EMBORCACAO	/ ITUMBIARA	C 1 MG	28
⊕ LT 500 kV GURUPI	/ MIRACEMA	C 2 TO	27
⊕ LT 500 kV SERRA MESA 2	/ LUZANIA	C 1 GO	25
⊕ LT 500 kV S.DA MESA	/ RIO DAS EGUAS	C 1 GO/BA	25
⊕ LT 500 kV IMPERATRIZ	/ COLINAS	C 2 MA/TO	25
⊕ LT 500 kV PEIXE 2	/ SERRA MESA 2	C 1 TO/GO	22
⊕ LT 500 kV ANGRA FUR	/ C. PAULISTA	C 1 RJ/SP	20
⊕ LT 500 kV EMBORCACAO	/ SAO GOTARDO 2	C 1 MG	19
⊕ LT 500 kV C. PAULISTA	/ CAMPINAS	C 1 SP	19
⊕ LT 500 kV JAURU	/ SUB CUIABA	C 1 MT	18

Com base nesses dados, foram analisados os 143 desligamentos da LT 500 kV GURUPI /MIRACEMA, localizada no estado do Tocantins, aos quais nos 116 desligamentos não haviam a detecção de raios, representando 80% da amostra analisada.

Em seguida, foram analisados os 114 desligamentos da linha LT 500 kV SERRA DA MESA /SAMAMBAIA, localizada na região Centro-oeste do país, aos quais 71 eventos não possuíam eventos de descargas, representando 62% de desligamentos descartados para essa análise.

A linha de transmissão LT 500 kV SERRA DA MESA / GURUPI, também localizada na região Centro-oeste do país foi analisada, possuindo 91 desligamentos por descargas atmosféricas. Porém, da mesma forma que as

demais linhas observadas acima, em 58 registros não havia a detecção de raios, representando 65% dessa amostra.

Prosseguindo com a análise das linhas de transmissão do setor de 500 kV, foi observada a LT 500 kV ANGRA FURNAS / CACHOEIRA PAULISTA, localizada nos estados do Rio de Janeiro e São Paulo. Foi observado que, ao seguir para outra região do país, foi obtido um comportamento diferente das demais linhas citadas, onde dos 20 eventos de desligamentos por descargas atmosféricas, apenas 3 não haviam detecção de raios, representando 15% dessa análise. Mesmo havendo uma evolução na correlação de desligamentos de linhas de transmissão e descargas atmosféricas detectadas, não havia uma frequência entre os anos de 2002 a 2015 representativa para seleção dessa linha, sendo necessária uma maior consistência anual desses dados para a definição adequada do equipamento a ser analisado nesse trabalho, pois 17 cenários de desligamentos não são suficientes para um adequado treinamento de uma Rede Neural MLP.

Em seguida foram analisadas as ocorrências associadas ao setor de 440kV, contendo 284 registros de desligamentos de linhas de transmissão. O estudo indicou uma grande abrangência dos desligamentos localizados na região de São Paulo, conforme apresentado na Tabela 5.4.

Tabela 5.4 - Quantidade de desligamentos no setor 440 kV

440	284
ND	284
⊕ LT 440 kV TAQUARUCU / JUPIA C 1 SP	32
⊕ LT 440 kV P. PRIMAVERA / TAQUARUCU C 1 SP	19
⊕ LT 440 kV P. PRIMAVERA / TAQUARUCU C 2 SP	17
⊕ LT 440 kV MOGI MIRIM 3 / SANTO ANGELO C 1 SP	17
⊕ LT 440 kV BAURU / CABREUVA C 1 SP	14
⊕ LT 440 kV RIBEIRAOPRETO / SANTA BARBARA C 1 SP	10
⊕ LT 440 kV ILHA SOLTEIRA / BAURU C 1 SP	10
⊕ LT 440 kV ILHA SOLTEIRA / BAURU C 2 SP	10
⊕ LT 440 kV OESTE / BAURU C 1 SP	9
⊕ LT 440 kV AGUA VERMELHA / RIBEIRAOPRETO C 1 SP	8
⊕ LT 440 kV BOM JARDIM / TAUBATE C 1 SP	8
⊕ LT 440 kV AGUA VERMELHA / ARARAQUARA CTP C 1 SP	7
⊕ LT 440 kV ILHA SOLTEIRA / AGUA VERMELHA C 1 SP	7
⊕ LT 440 kV OESTE / BAURU C 2 SP	7
⊕ LT 440 kV JUPIA / BAURU C 2 SP	6
⊕ LT 440 kV ARARAQUARA CTP / SANTO ANGELO C 1 SP	6

A linha de transmissão LT 440 kV TAQUARUÇU / JUPIA, com 32 registros de desligamentos foi analisada, observando uma distribuição uniforme das ocorrências dos mesmos durante o período de 2002 e 2015, tornando a análise mais consistente. Dos 32 desligamentos ocorridos, em apenas 1 deles não houve a detecção de descargas atmosféricas próximas a linha. A boa observabilidade dos registros de incidência de raios durante os períodos de desligamento dessa linha de transmissão, indicou como uma linha válida para prosseguimento do estudo, contendo 31 cenas com desligamento para treinamento do modelo RDLR.

Diante da análise dessa etapa, foram evidenciados que outros fatores devem ser considerados na seleção adequada da linha. O primeiro, refere-se a escolha de uma região do mapa de densidade de descargas atmosféricas com um nível cerâmico relevante, conforme representado na Figura 2.4 do Capítulo 2. O segundo fator importante, é que se trata de uma região do país com boa eficiência da rede de detecção de descargas atmosféricas, conforme indicado na Figura 2.3 do Capítulo 2, para garantir uma correlação verdadeira entre a incidência de descargas e desligamentos de linhas de transmissão.

Dessa forma, a escolha da linha de transmissão linha LT 440kV TAQUARUÇU / JUPIÁ foi definida para garantir uma melhor qualidade na detecção de descargas atmosféricas, assegurada por uma região com uma relevante incidência de raios, conforme apresentado na Tabela 5.5.

Tabela 5.5 – Critério de Seleção da Linha de Transmissão

Linha de Transmissão	Desligamentos	Desligamentos sem Incidência de Raios	Desligamentos sem Incidência de Raios (%)
LT 500 kV GURUPI /MIRACEMA	143	116	81
LT 500 kV SERRA DA MESA /SAMAMBAIA	114	71	62
LT 500 kV SERRA DA MESA / GURUPI	89	58	65
LT 500 kV ANGRA FURNAS / CACHOEIRA PAULISTA	20	3	15
LT 440 kV TAQUARUÇU / JUPIA	32	1	3

5.2.2.

Critério para Definição do Espaço Amostral de Descargas Atmosféricas

Após a seleção adequada de uma linha de transmissão para treinamento do modelo RDLR, é importante definir o espaço amostral de descargas atmosféricas, que serão posteriormente agrupadas para a observação e estudo do padrão desses eventos no desligamento de linhas de transmissão.

Para tal, foi necessário definir um critério para identificação desse espaço amostral, observando eventos de descargas atmosféricas a 1.000 km ($d_{ce_{MAX}}$) da linha de transmissão LT 440 kV TAQUARUÇU / JUPIÁ, ocorridas no intervalo de 3 horas a 30 minutos antes do momento dos desligamentos da linha, o que denominamos “cenas com desligamentos”. Partindo desse critério, a amostra resultou em 499.369 registros, contendo informações que permitiram relacionar cada raio aos respectivos 31 desligamentos (cenas), com causa determinada por descargas atmosféricas, compreendidos no período de 2002 a 2015.

Cada registro possui os atributos dos dados de descargas atmosféricas, descritos no item 3.1, que serão utilizadas como variáveis de entrada na Rede Neural Multilayer Perceptron (MLP), no modelo de determinação do risco de desligamento de linhas de transmissão nesse trabalho.

A Figura 5.2 representa geograficamente o espaço amostral de descargas atmosféricas em torno da linha LT 440 kV TAQUARUÇU / JUIPIA, onde cada cor na figura representa o conjunto de descargas para cada um dos desligamentos registrados para essa linha de transmissão. Na legenda da figura observamos cada desligamento ocorrido e a quantidade de descargas observadas dentro intervalo anterior a cada ocorrência, entre colchetes.

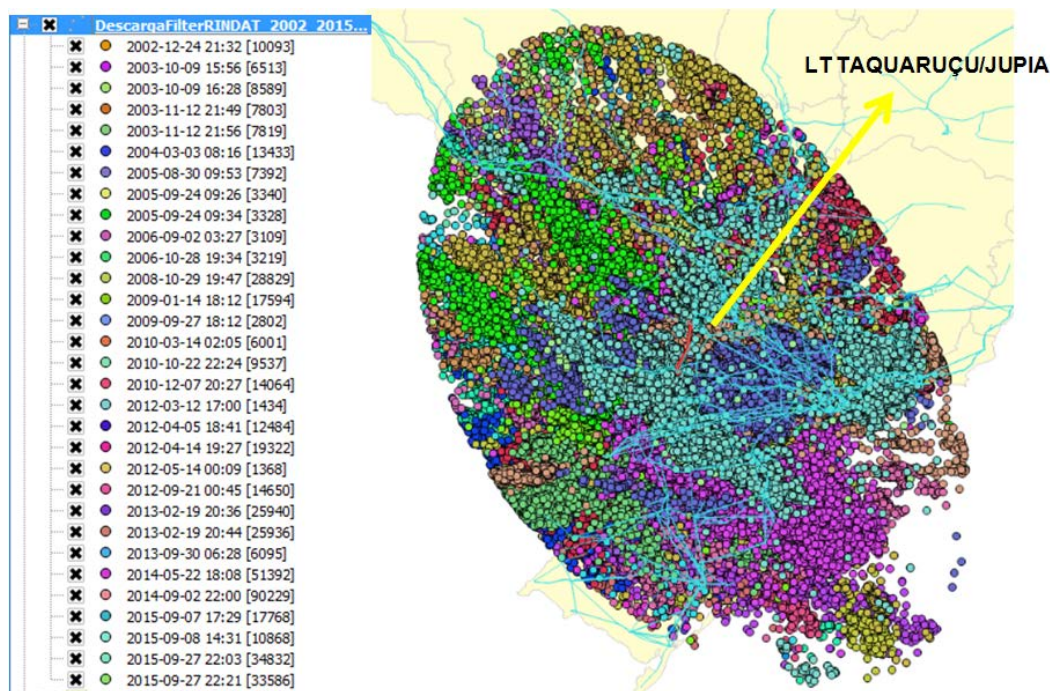


Figura 5.2 - Espaço amostral geográfico de descargas atmosféricas em torno da LT 440 kV TAQUARUÇU/JUPIÁ associados a desligamentos da linha de transmissão

Além do espaço amostral das descargas atmosféricas associadas aos desligamentos da linha de transmissão LT 440 kV TAQUARUÇU/JUPIÁ, o trabalho considerou também o espaço amostral de descargas atmosféricas próximas a essa linha de transmissão, em momentos que não ocorreram desligamentos, o que denominamos de “cenas sem desligamento”. A partir desse conceito, foi realizada uma consulta aleatória ao banco de dados de descargas atmosféricas, onde foram observadas 275 cenas contendo raios a pelo menos 1.000 km da LT 440 kV TAQUARUÇU/JUPIÁ, compreendidos entre 2002 e 2015, associados aos instantes de tempo aleatórios produzidos. Para definição das cenas mais relevantes que não ocorreram desligamentos para o estudo, foi realizada uma ordenação decrescente por quantidade de descargas, para a seleção das 31 primeiras cenas, configurando a mesma quantidade de cenas que ocorreram desligamentos.

Desta forma, foi possível produzir uma amostra com a mesma quantidade de cenas (31) sem desligamento, com um conjunto elevado de raios. Essa amostra é representada por 283.365 registros de descargas atmosféricas, ocorridas a 1.000 km da LT 440 kV TAQUARUÇU/JUPIÁ, em momentos que não

ocorreram desligamentos, produzindo 31 cenas que variam de 4.168 a 28.520 raios por cena, conforme apresentado na Tabela 5.6.

**Tabela 5.6 – Cenas sem desligamentos da LT 440 kV
TAQUARUÇU/JUPIÁ ordenadas pela quantidade de raios**

Cena	Data/Hora	Quantidade de Raios
1	'2014-11-07 03:00:52'	28520
2	'2004-12-07 20:01:43'	15032
3	'2013-03-15 21:45:14'	14817
4	'2012-10-28 22:33:03'	14276
5	'2013-06-01 20:33:57'	12989
6	'2013-09-22 23:20:33'	11832
7	'2012-12-30 00:02:52'	11721
8	'2010-01-26 17:10:54'	11559
9	'2010-04-23 07:11:36'	10142
10	'2009-05-14 15:14:39'	9844
11	'2013-02-02 16:10:42'	9842
12	'2012-06-04 03:03:39'	8426
13	'2009-01-30 19:49:41'	8274
14	'2012-04-14 11:05:38'	8264
15	'2015-03-29 21:43:39'	7997
16	'2015-02-14 00:09:56'	7784
17	'2013-04-09 22:31:52'	7688
18	'2006-12-26 18:46:37'	7662
19	'2010-03-29 18:00:59'	7639
20	'2008-11-10 00:15:32'	7338
21	'2014-04-12 04:56:58'	7225
22	'2014-09-18 09:55:19'	7018
23	'2009-02-23 09:23:24'	6617
24	'2007-10-22 00:56:36'	6328
25	'2014-01-14 18:37:36'	5901
26	'2014-02-14 03:28:41'	5649
27	'2009-01-26 19:53:36'	5276
28	'2010-12-22 18:47:40'	5003
29	'2008-11-15 22:15:24'	4336
30	'2003-04-10 07:49:16'	4198
31	'2013-02-02 05:14:10'	4168
TOTAL		283365

5.3.

Agrupamento do espaço amostral de descargas atmosféricas

O algoritmo DBSCAN (*Density Based Spatial Clustering of Application with Noise* (*Clusterização Espacial Baseada em Densidade de Aplicações com Ruído*)) foi utilizado nesse estudo como o método de agrupamento dos dados, por ser capaz de identificar grupos de tamanhos e formatos diferentes, bem como separação eficiente de ruídos (*outliers*).

O algoritmo DBSCAN foi testado para descargas atmosféricas próximas a linhas de transmissão, apresentando bom desempenho no agrupamento dos eventos de descarga e na identificação de ruídos, sendo possível observar a sensibilidade no ajuste dos parâmetros de entrada Eps (raio de vizinhança dos elementos do grupo) e MinPts (número mínimo de elementos por grupo). O estudo foi realizado sobre uma amostra de descargas atmosférica no dia 26/09/2014 às 04hs01min, próximas a linha de transmissão LT Cascavel Oeste/Ivaiporã C1, para diferentes valores de Eps e MinPts, observadas em intervalos a cada 30 minutos. No eixo x foi representada a longitude e no eixo y a latitude, sendo os agrupamentos diferenciados através de cores. Os resultados completos do estudo para os diferentes valores dos parâmetros Eps e MinPts estão apresentados no Anexo III. A Tabela 12 apresenta a descrição dos parâmetros do modelo.

A análise dos resultados produzidos pelo algoritmo DBSCAN no agrupamento dos dados de descargas atmosféricas permitiu observar que para valores altos dos parâmetros Eps e MinPts, foram gerados poucos grupos, porém muito densos. Por outro lado, para valores baixos de Eps e MinPts foram gerados muitos grupos e menos densos. No teste realizado com alto valor de Eps e baixo para MinPts foram observados grandes agrupamentos, porém menos densos. Para valores baixos de Eps e altos de MinPts foram gerados pequenos grupos muito densos. Além dessas observações, valores pequenos de Eps produzem muitos ruídos, podendo eliminar muitos dados relevantes, conforme é possível observar na Figura 5.3. A Tabela 5.7 resume os resultados observados nesse estudo e os valores utilizados nas análises estão na Tabela 5.8.

Tabela 5.7 – Resultados do agrupamento com algoritmos DBSCAN

Valores de Eps	Valores de MinPts	Resultado
Alto ↑	Alto ↑	Poucos clusters grandes e densos
Baixo ↓	Baixo ↓	Muitos clusters pequenos e menos densos
Alto ↑	Baixo ↓	Clusters grandes de menos densos
Baixo ↓	Alto ↑	Clusters pequenos e densos

Tabela 5.8 – Valores dos parâmetros Eps e MinPts analisados no estudo

Valores de Eps	Valores de MinPts
0,33588	5
0,26283	5
0,34688	5
0,50054	5
0,45758	5
0,4169	5
0,51598	5
0,656	5
0,35113	5
0,2	5
0,05	5

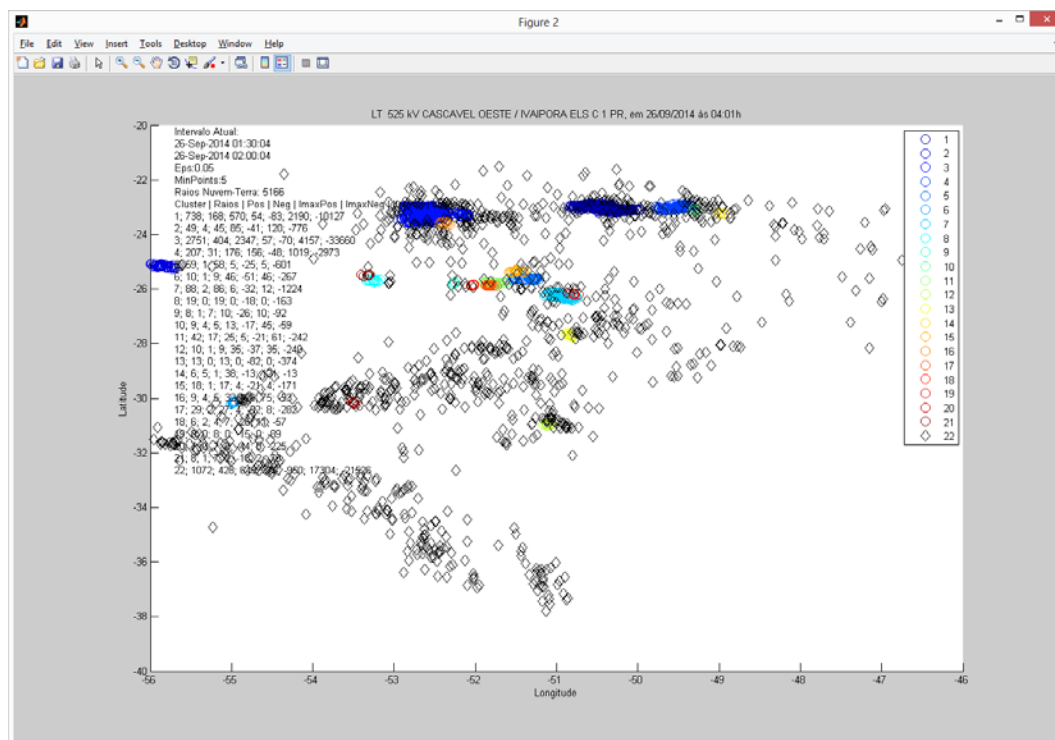


Figura 5.3 - DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 02:00:04 a 02:30:04)

Dessa forma, os parâmetros de agrupamento utilizados no método foram raio de vizinhança Eps=0,33 e número mínimo de pontos MinPts=5, que levaram aos melhores resultados no processo de agrupamento, conforme é observado na Figura 5.4.

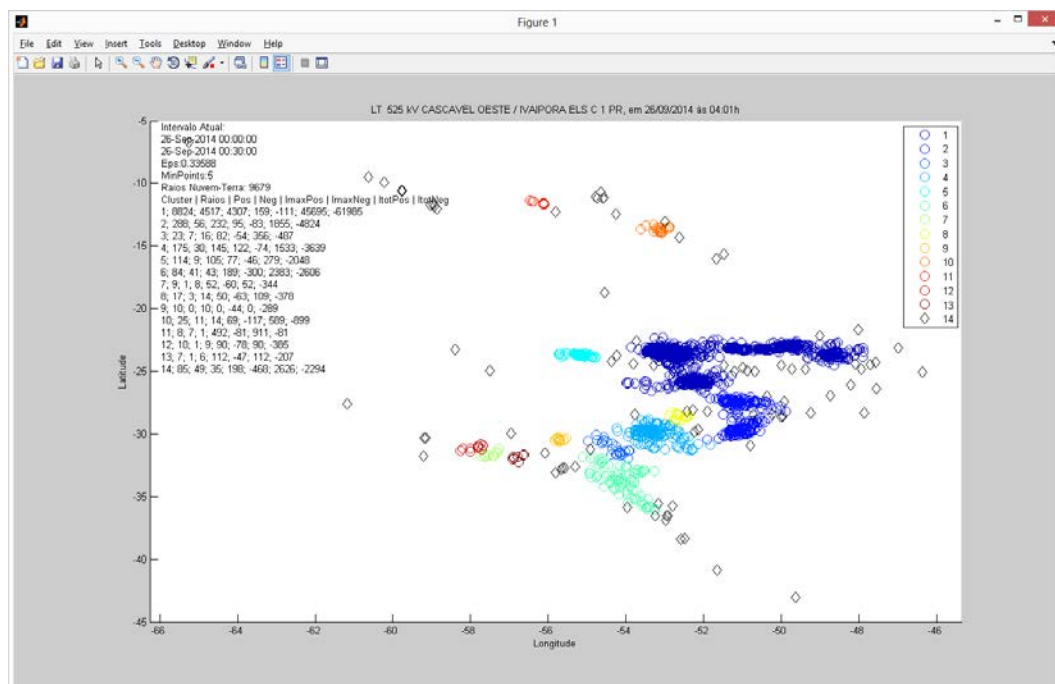


Figura 5.4 – DBSCAN para Eps=0,33 e MinPts=5 (Intervalo 00:00:00 a 00:30:00)

Para os 31 desligamentos associados LT 440 kV TAQUARUÇU / JUPIA, foram gerados grupos através do método de agrupamento baseado em densidade (DBSCAN), detalhado no Anexo III. A Figura 5.5 representa a distribuição dos grupos de descargas atmosféricas, referentes ao cenário de desligamento da linha de transmissão selecionada nesse estudo, ocorrido no dia 09 de outubro de 2003 às 16:28:00 horas.

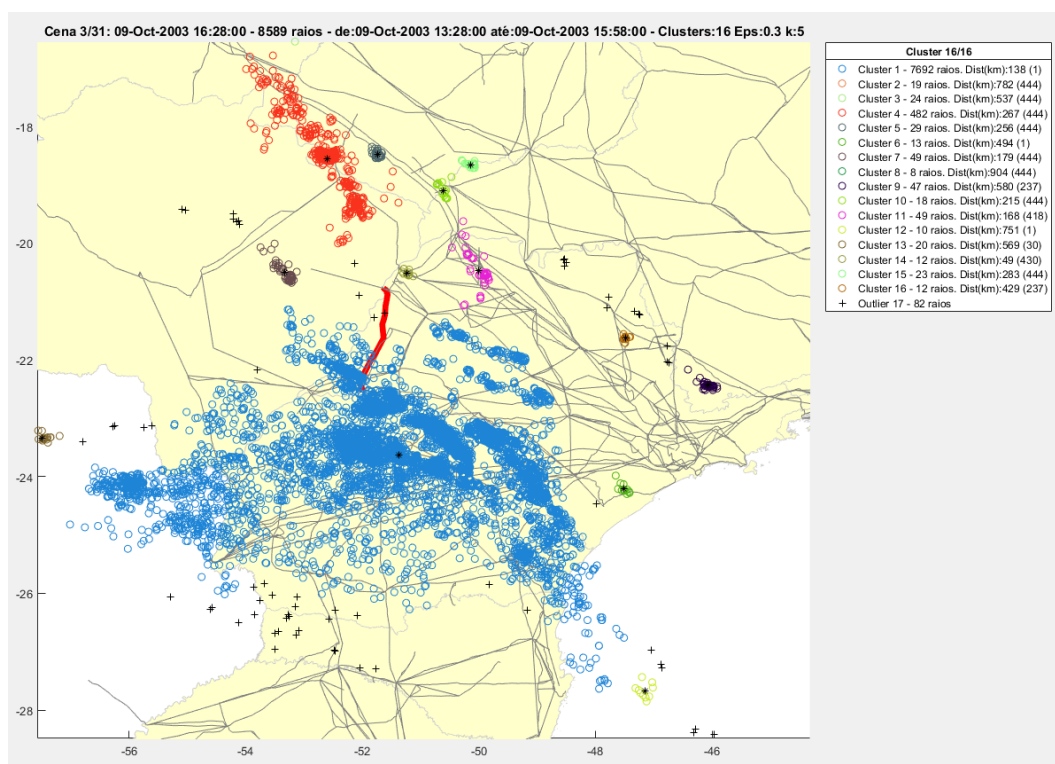


Figura 5.5 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 3

A mesma situação se aplica aos 31 registros de descargas atmosféricas em torno da LT 440 kV TAQUARUÇU / JUPIA, em momentos que não ocorreram desligamentos da linha de transmissão, conforme ilustradas nas Figuras 5.6, 5.7 e 5.8.

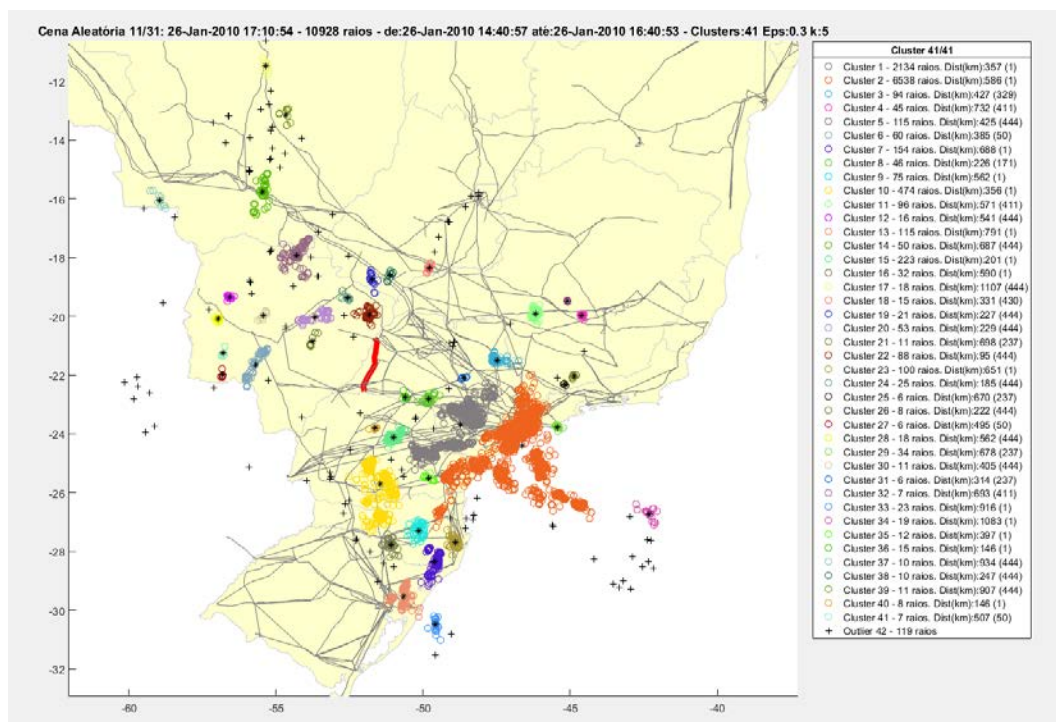


Figura 5.6 - Agrupamento das descargas atmosféricas sem a ocorrência de desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 11

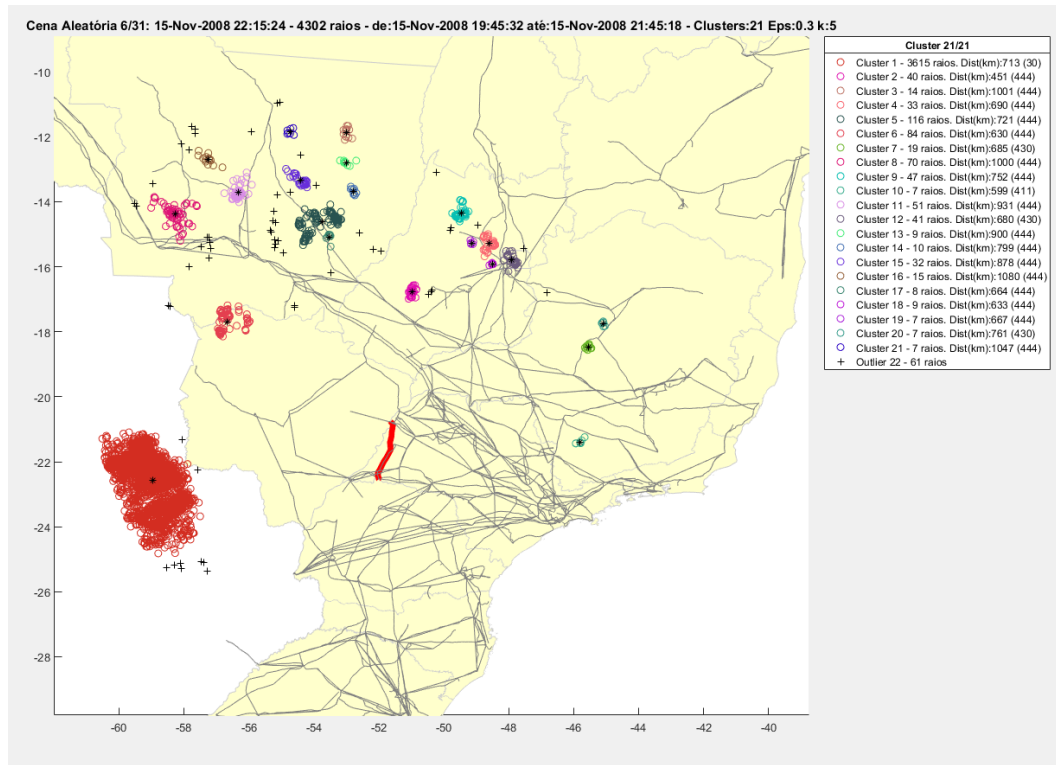


Figura 5.7 - Agrupamento das descargas atmosféricas sem a ocorrência de desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 6

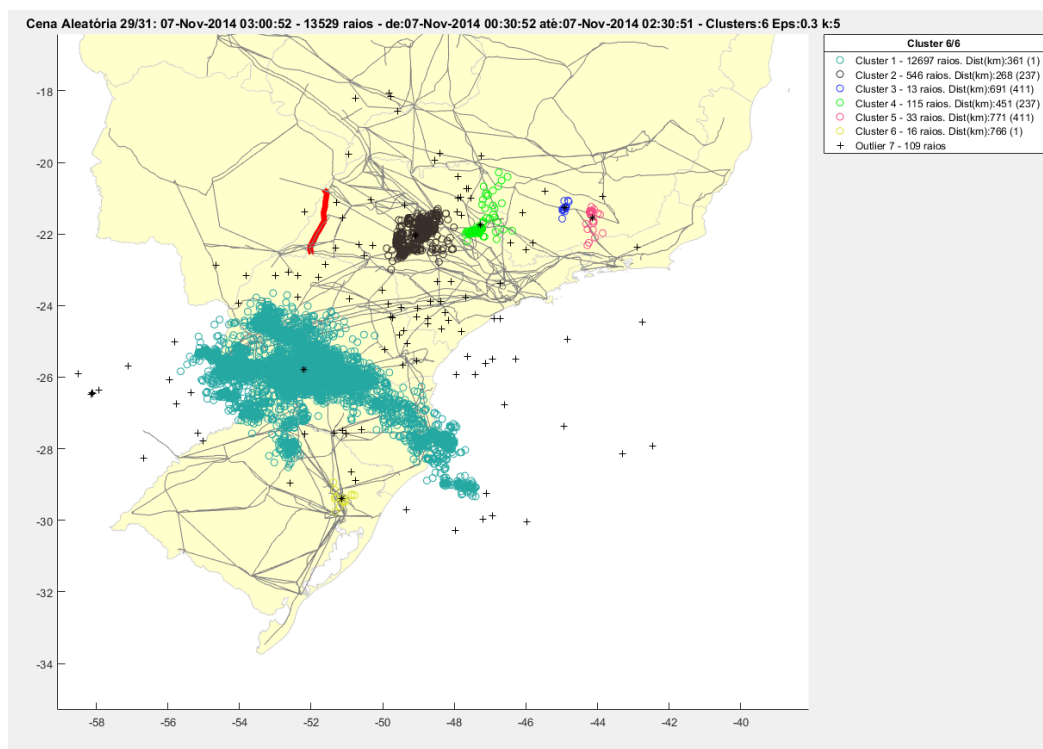


Figura 5.8 - Agrupamento das descargas atmosféricas sem a ocorrência de desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 29

5.4.

O Modelo da Rede Neural *Multilayer Perceptron* (MLP)

A previsão do risco de desligamento de linhas de transmissão por descargas atmosféricas é baseada numa Rede Neural *Multilayer Perceptron* (MLP). O estudo de caso se refere aos desligamentos ocorridos na linha de transmissão LT TAQUARUÇU/JUPIA, identificados pela causa associada a descargas atmosféricas, conforme descritos no Capítulo 3 e uma amostra de 782.734 registros de descargas atmosféricas ocorridas à 1.000km da linha, distribuídas em 62 cenas (31 em que ocorreram desligamentos e 31 sem a ocorrência de desligamento), conforme análise dos dados e critérios de pré-processamento descritos nos itens 5.1 e 5.2.

5.4.1.

Variáveis de Entrada dos Agrupamentos de Descargas Atmosféricas

Cada descarga atmosférica que compõe os agrupamentos possui informações individuais que podem caracterizar efetivamente os seus grupos. As variáveis de entrada dos agrupamentos de descargas atmosféricas são definidas como: a distância do centroide do grupo à linha de transmissão, o total de raios do grupo, número de descargas com polaridade positiva, número de descargas com polaridade negativa, corrente máxima positiva e negativa do grupo, corrente total positiva e negativa, corrente média e mediana, tempo de ascensão médio e mediano, tempo de pico a zero médio e mediano do grupo.

Além dessas variáveis de entrada, são consideradas também a sazonalidade e a quantidade de raios do evento, totalizando 16 variáveis de entrada por cada grupo a ser processado na Rede Neural MLP. As variáveis de entrada representam 44 entradas para cada uma das 31 amostras de desligamentos da linha de transmissão sob análise e 31 registros de amostras com descargas atmosféricas próximas a linha de transmissão, sem haver a ocorrência de desligamento.

Nesse estudo foi considerado o treinamento dos 3 primeiros grupos mais representativos na identificação do padrão de raios responsável pelo desligamento de linhas de transmissão. Para a identificação dos 3 grupos mais representativos foram elaborados 3 métodos, conforme distância do centroide do

agrupamento à linha de transmissão, quantidade de raios e razão entre quantidade de raios e distância do centroide à linha de transmissão. A seguir, são descritos esses 3 métodos com a identificação daquele que apresentou melhores resultados na identificação dos grupos mais relevantes.

5.4.2.

Seleção dos Agrupamentos Relevantes

Após o agrupamento do espaço amostral de descargas atmosféricas definido nesse estudo, foram determinados critérios para seleção dos 3 grupos mais relevantes para o treinamento da Rede Neural MLP, de forma a garantir melhores resultados e maior desempenho da rede. Dessa forma, foram definidos 3 diferentes critérios, sendo o primeiro baseado na distância do centroide de cada grupo à linha de transmissão, o segundo baseado na quantidade de raios dos agrupamentos e o último baseado na razão entre quantidade de raios e distância do centroide à linha de transmissão.

5.4.2.1.

Agrupamentos selecionados pelo critério da distância do grupo à Linha de Transmissão (Critério 1)

Esse critério é baseado na seleção dos agrupamentos conforme a distância do centroide à linha de transmissão para a definição dos grupos, sendo os mais relevantes aqueles mais próximos à linha. Através da análise dos resultados obtidos da matriz de confusão, é possível observar que a Rede Neural MLP obteve 80,6% de sucesso na detecção dos desligamentos da linha de transmissão por descargas atmosféricas e 83,9% sucesso na identificação das amostras de descargas atmosféricas que não causaram desligamentos da linha de transmissão. Nas amostras de descargas atmosféricas que causaram desligamentos da linha de transmissão, representando um total de 50% da amostra, foram observados 40,3% de resultados positivos e 9,7% de resultados falso positivos. Para as amostras que não causaram desligamentos da linha de transmissão, representando também 50% da amostra, foram observados 41,9% de resultados positivos e 8,1% de resultados falso positivo. O resultado final da matriz de confusão foi de 82,3%, conforme Figura 5.9.

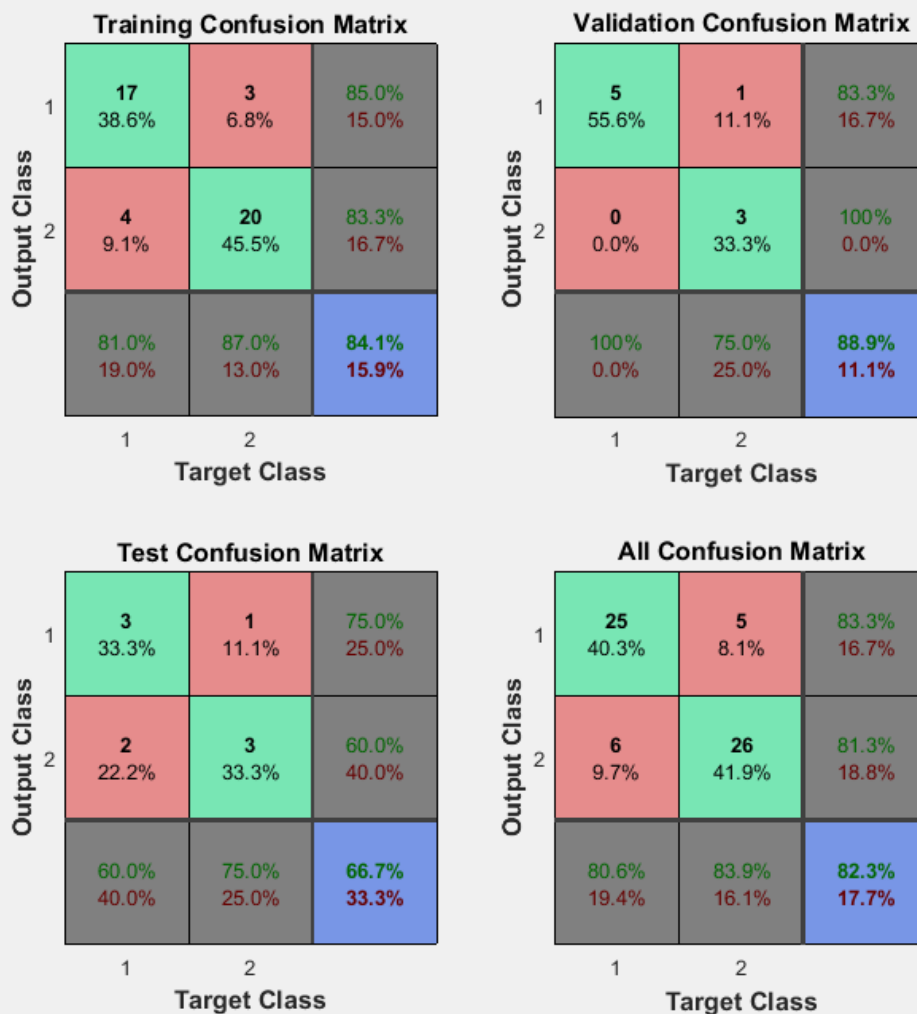


Figura 5.9 – Matriz de Confusão – Agrupamentos ordenados pela distância do centroide do grupo à LT

O gráfico representado na Figura 5.10 apresentada o desempenho da rede neural para o conjunto de treinamento, validação e teste. É possível observar que o melhor desempenho foi obtido na época 19.

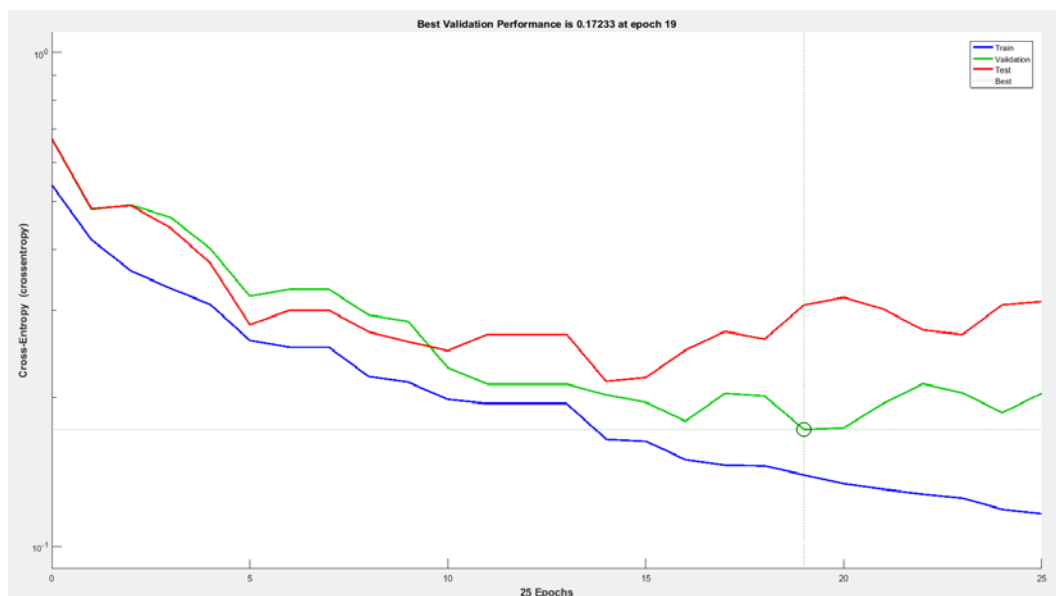


Figura 5.10 – Desempenho da Rede Neural – Agrupamentos ordenados pela distância do centroide do grupo à LT

5.4.2.2.

Agrupamentos selecionados pelo critério da quantidade de raios (Critério 2)

Esse critério é baseado na seleção dos agrupamentos conforme a quantidade de raios de cada grupo para a definição daqueles mais relevantes, sendo os mais relevantes os agrupamentos mais numerosos. Através da análise dos resultados obtidos da matriz de confusão, é possível observar que a Rede Neural MLP obteve 80,6% de sucesso na detecção dos desligamentos da linha de transmissão por descargas atmosféricas e 90,3% sucesso na identificação das amostras de descargas atmosféricas quando não ocorreram desligamentos da linha de transmissão, representando 50% do total da amostra. Nas amostras de descargas atmosféricas que houveram desligamentos da linha de transmissão, também 50% do total, foram observados 40,3% de resultados positivos e 9,7% de resultados falso positivos. Para as amostras sem a ocorrência de desligamentos da linha de transmissão, foram observados 45,2% de resultados positivos e 4,8% de resultados falso positivos. O resultado final da matriz de confusão foi de 85,5%, conforme Figura 5.11.

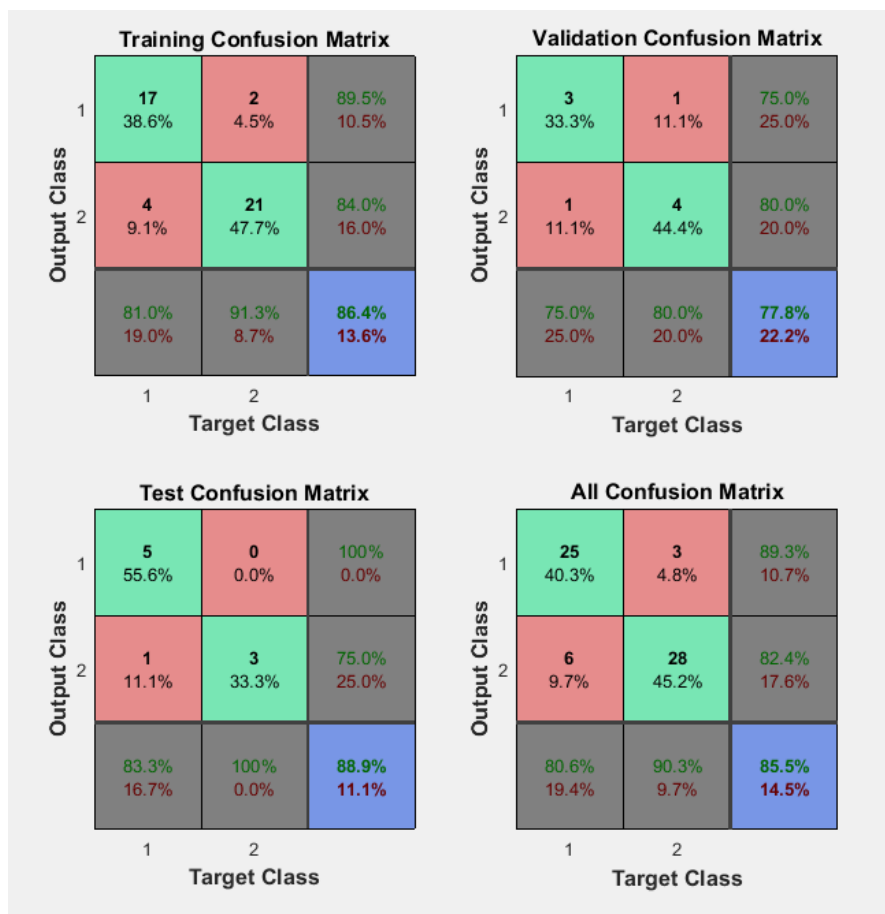


Figura 5.11 – Matriz de Confusão – Agrupamentos ordenados pela quantidade de raios

O gráfico representado na Figura 5.12 apresentada o desempenho da rede neural para o conjunto de treinamento, validação e teste. É indicado o melhor desempenho na época 5.

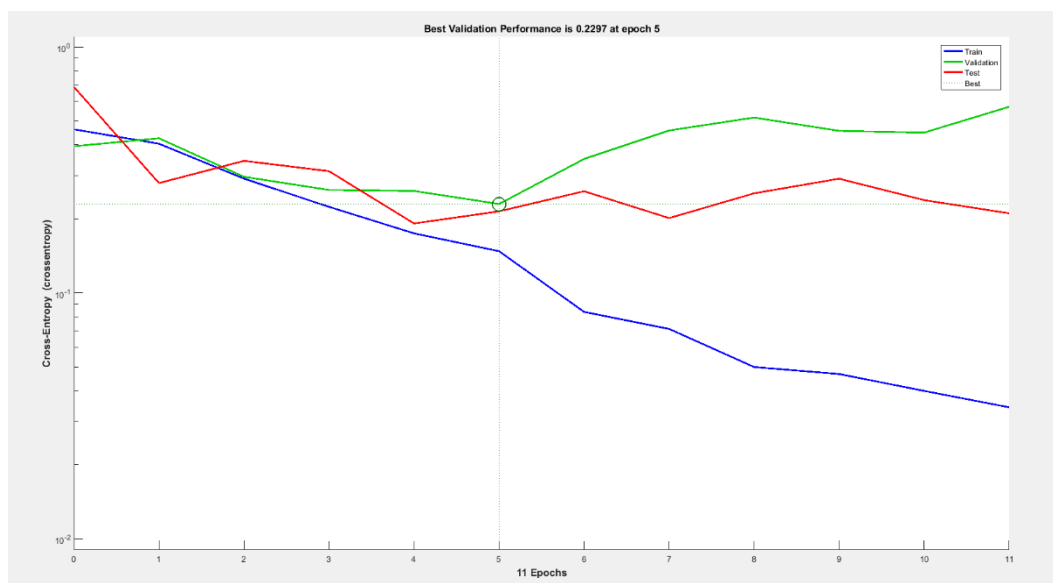


Figura 5.12 – Desempenho da Rede Neural – Clusters ordenados pela quantidade de raios

5.4.2.3.

Agrupamentos selecionados pelo critério da razão entre quantidade de raios e distância do centroide à linha de transmissão (Critério 3)

Foi analisado um último critério para seleção dos agrupamentos mais relevantes, considerando a razão entre quantidade de raios e distância do centroide à linha de transmissão. Através da análise dos resultados obtidos da matriz de confusão, é possível observar que a Rede Neural MLP obteve 90,3% de sucesso na detecção dos desligamentos da linha de transmissão por descargas atmosféricas e 100% de sucesso na identificação das amostras de descargas atmosféricas quando não ocorreram desligamentos da linha de transmissão. Nas amostras de descargas atmosféricas associadas aos desligamentos da linha de transmissão (50% do total), foram observados 45,2% de resultados positivos e 4,8% de resultados falso positivos. Para as amostras relacionadas aos eventos que não ocorreram desligamentos da linha de transmissão, foram observados 50% de resultados positivos e nenhum resultado falso positivo, também para 50% da amostra. O resultado final da matriz de confusão foi de 95,2%, conforme apresentado na Figura 5.13.



Figura 5.13 – Matriz de Confusão – Agrupamentos ordenados pela razão da quantidade de raios e a distância do centroide à LT

O gráfico representado na Figura 5.14 apresentada o desempenho da rede neural para o conjunto de treinamento, validação e teste, sendo observado o melhor desempenho na época 16.

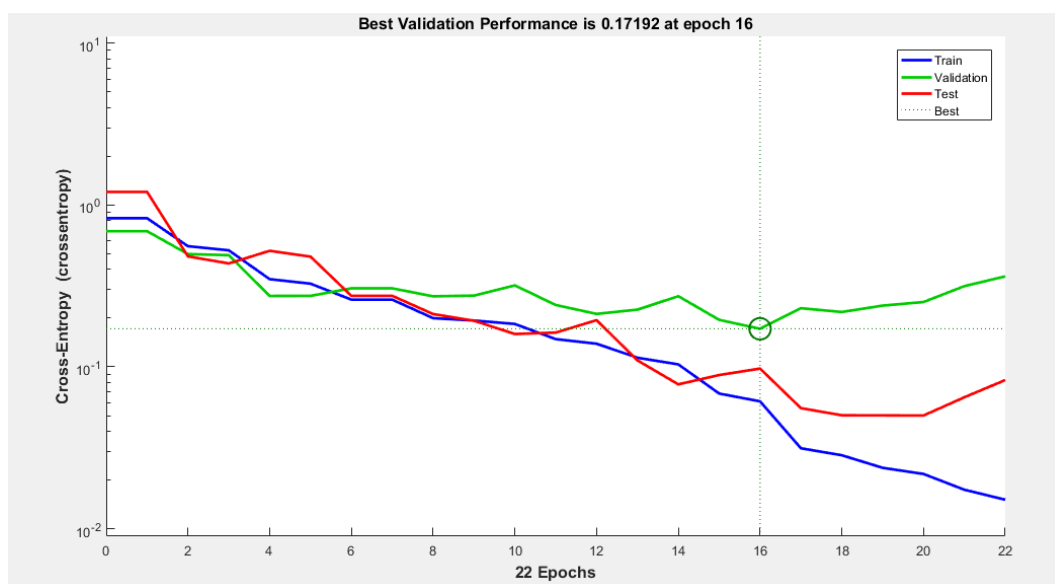


Figura 5.14 – Desempenho da Rede Neural – Agrupamentos ordenados pela razão da quantidade de raios e distância do centroide à LT

O resultado da análise sobre os 3 critérios para a seleção dos agrupamentos e mais relevantes para o modelo RDLR desse estudo, indicou que o critério dos grupos selecionados conforme razão da quantidade de raios e distância do centroide à linha de transmissão apresentou melhor resultado avaliado no desempenho desse modelo, sendo definido como o critério de seleção de agrupamentos nesse trabalho. A Tabela 5.9 reflete os resultados obtidos da análise comparativa dos 3 métodos propostos nesse estudo.

Tabela 5.9 – Análise comparativa dos métodos de seleção dos agrupamentos

Métodos de seleção dos agrupamentos	Agrupamentos associados ao desligamento da LT	Agrupamentos associados a não ocorrência de desligamento da LT
Critério 1	80,60%	90,30%
	Desempenho resultante da Rede Neural = 82,3%	
Critério 2	80,60%	83,90%
	Desempenho resultante da Rede Neural = 85,5%	
Critério 3	90,30%	100,00%
	Desempenho resultante da Rede Neural = 95,2%	

5.4.3. Seleção das Variáveis de Entrada

O elevado grau de complexidade presente nas Redes Neurais Artificiais pode prejudicar seu desempenho, e a presença de variáveis de entrada redundantes, que nada acrescentam ao seu processo de aprendizagem, tornam o treinamento mais difícil e demorado. Existem vários métodos com o objetivo de selecionar quais variáveis de entrada possuem características mais relevantes para a determinação da saída ou resposta da rede, possibilitando a redução do número de entradas.

Nesse estudo foi aplicado o método de Análise de Componentes Principais, conhecida como *Principal Components Analysis* (PCA), que é uma técnica estatística utilizada para reduzir a dimensionalidade de um conjunto de dados onde há um grande número de variáveis inter-relacionadas. Essa redução é realizada devido a um novo e menor conjunto de variáveis não correlacionadas, chamadas componentes principais [30]. Dessa forma, as componentes principais selecionadas permitem reduzir a dimensão do conjunto de dados e não o número de variáveis originais, já que são combinações lineares de todas variáveis originais.

Fazer uma seleção de quais informações serão relevantes como variáveis de entrada para a determinação do risco de desligamento de linhas de transmissão é algo bastante complexo, devido a quantidade de atributos associados às descargas atmosféricas que compõem os agrupamentos. Portanto, deve-se fazer uma análise minuciosa sobre quais dessas informações são mais importantes na representação do sistema.

Aplicando o método PCA nesse estudo, foi possível observar as variáveis explicadas para cada componente principal, conforme apresentado na Figura 5.15. Através desse gráfico, é possível verificar que o maior valor de variância explicada dos dados é representado pela primeira componente principal, a qual explica 38% da variância total dos dados originais. As três primeiras componentes selecionadas para representação dos dados apresentam 97,6% da variância explicada acumulada. A partir da terceira componente principal, a variância acumulada começa a saturar e atinge 100% nas trinta primeiras componentes principais, a partir da qual não existirão componentes principais com relevância.

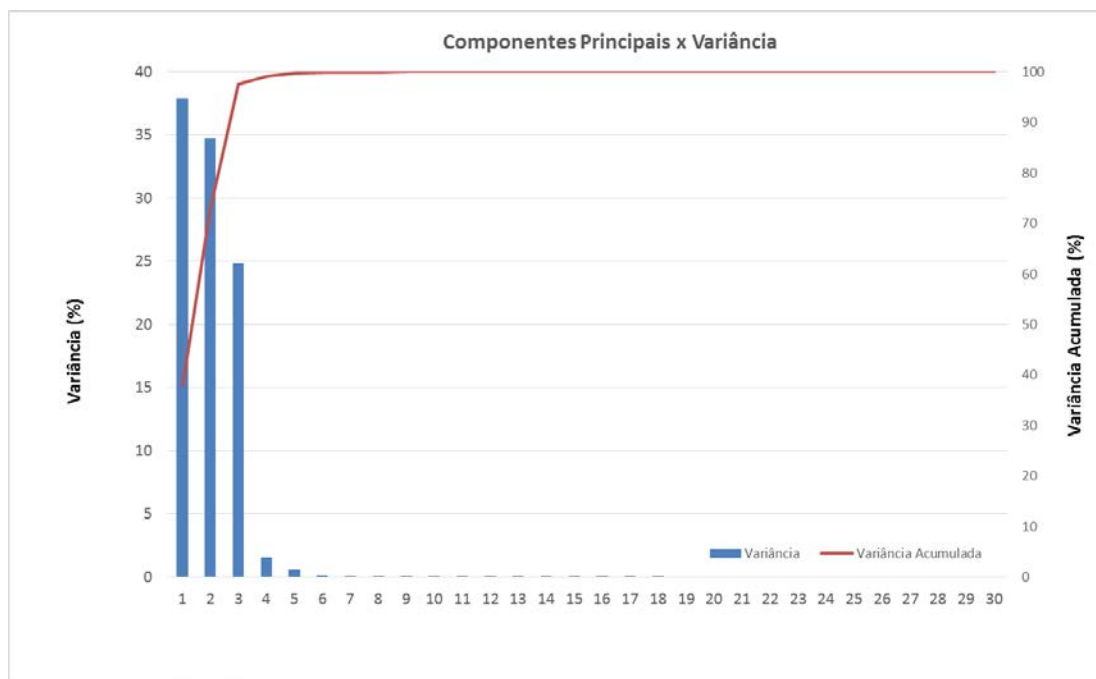


Figura 5.15 – Gráfico de barras das Componentes Principais x Variância

A Figura 5.16 representa os *scores* das entradas sobre as duas primeiras componentes principais e na Figura 5.17 é possível observar os *scores* das três primeiras componentes principais.

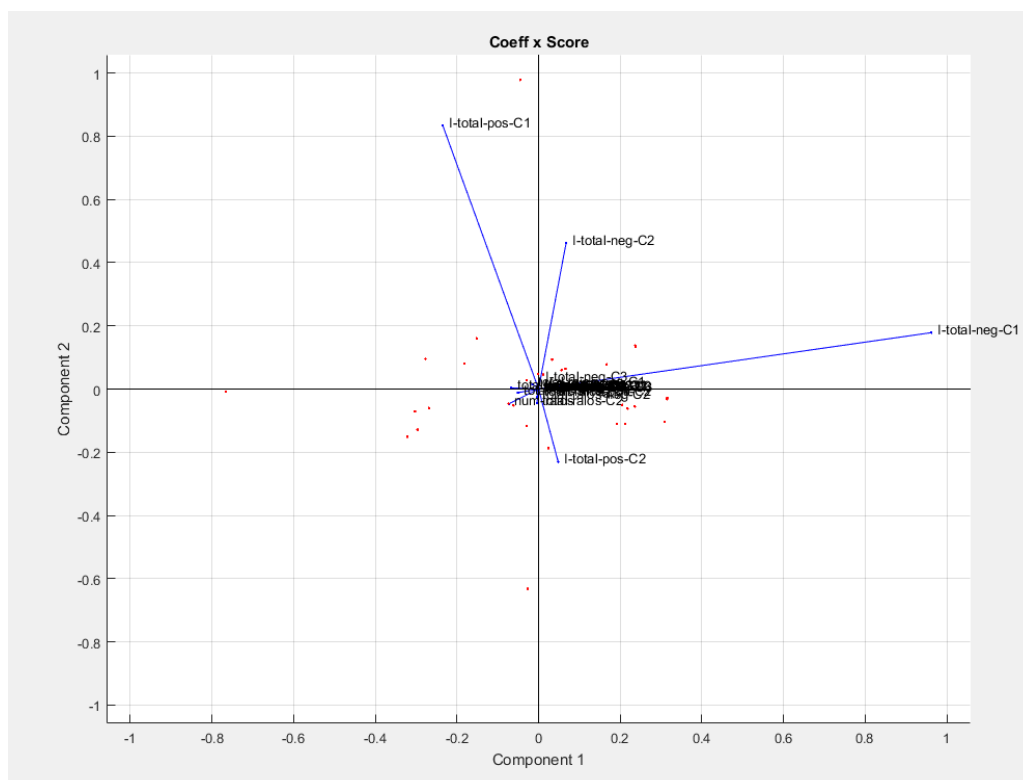


Figura 5.16 – Score das entradas sobre as duas primeiras componentes principais

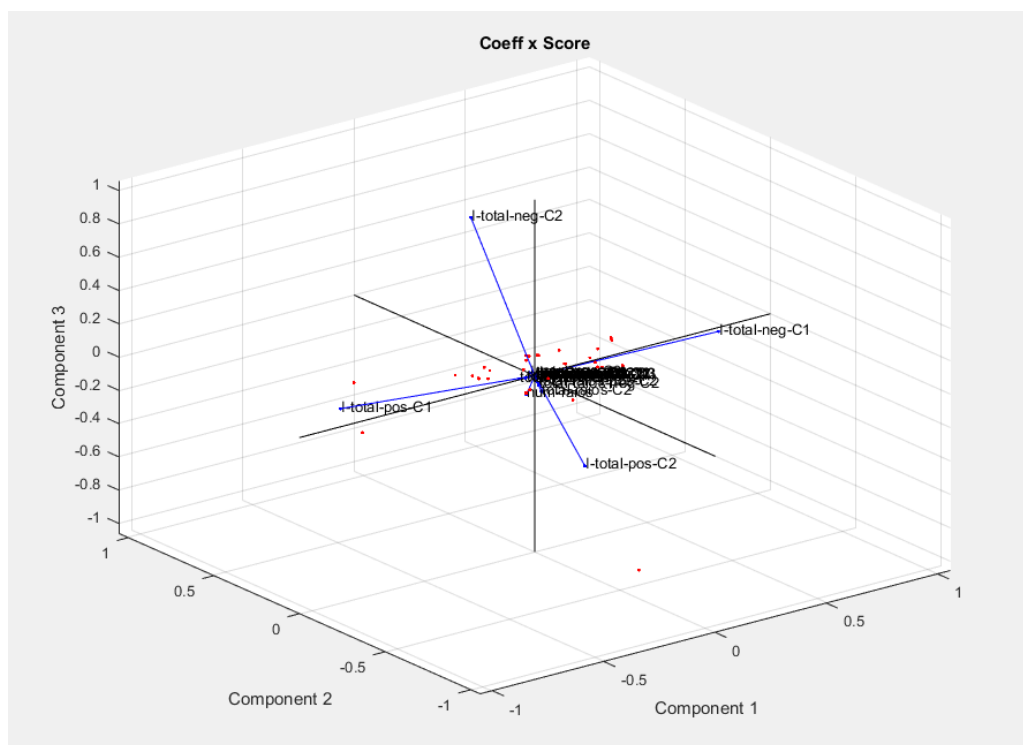


Figura 5.17 - Score das entradas sobre as três primeiras componentes principais

O método proposto avalia a seleção das variáveis baseado na raiz quadrada da soma do quadrado sobre cada componente principal, produzindo a indicação das variáveis de entrada, ordenadas de acordo com sua relevância para os componentes principais considerados. Para atingir 100% da variância acumulada, foram considerados os 30 componentes principais sobre as 44 variáveis de entrada. Esse método foi analisado comparativamente com o método PCA modificado, o qual é um algoritmo que também permite a seleção de variáveis. O método PCA modificado consiste em um algoritmo iterativo, onde a cada execução é selecionada a variável de entrada que possui maior importância e remove esse elemento da matriz de variáveis, prosseguindo assim para a próxima interação. Dessa forma, em cada processo iterativo é selecionada a variável mais representativa do componente principal, formando um conjunto de variáveis de entrada selecionadas conforme seu peso no componente principal.

O resultado dessa análise comparativa permitiu observar resultados muito semelhantes entre os dois métodos, sendo que a metodologia proposta nesse estudo apresentou um melhor desempenho por não ser um método iterativo. A

Tabela 5.10 representa a análise comparativa entre os métodos citados para a seleção de variáveis.

Tabela 5.10 - Resultados da análise comparativa entre o método proposto e PCA Modificado

Ran k	PCA 30		PCA 30 Modificado	
#	Entrada	Descrição	Entrada	Descrição
1	9	Corrente total positiva (I-total-pos-C1)	10	Corrente total negativa (I-total-neg-C1)
2	10	Corrente total negativa (I-total-neg-C1)	9	Corrente total positiva (I-total-pos-C1)
3	23	Corrente total positiva (I-total-pos-C2)	24	Corrente total negativa (I-total-neg-C2)
4	24	Corrente total negativa (I-total-neg-C2)	23	Corrente total positiva (I-total-pos-C2)
5	38	Corrente total negativa (I-total-neg-C3)	38	Corrente total negativa (I-total-neg-C3)
6	2	Número de raios (num-raios)	37	Corrente total positiva (I-total-pos-C3)
7	37	Corrente total positiva (I-total-pos-C3)	2	Número de raios (num-raios)
8	31	Distância agrupamento à LT (dist-C3)	5	Total de raios positivos (total-raios-pos-C1)
9	6	Total de raios negativos (total-raios-neg-C1)	18	Total de raios (total-raios-C2)
10	4	Total de raios (total-raios-C1)	6	Total de raios negativos (total-raios-neg-C1)
11	5	Total de raios positivos (total-raios-pos-C1)	4	Total de raios (total-raios-C1)
12	3	Distância agrupamento à LT (dist-C1)	32	Total de raios (total-raios-C3)
13	17	Distância agrupamento à LT (dist-C2)	20	Total de raios negativos (total-raios-neg-C2)
14	20	Total de raios negativos (total-raios-neg-C2)	31	Distância agrupamento à LT (dist-C3)
15	18	Total de raios (total-raios-C2)	17	Distância agrupamento à LT (dist-C2)
16	19	Total de raios positivos (total-raios-pos-C2)	3	Distância agrupamento à LT (dist-C1)
17	22	Corrente máxima negativa (I-max-neg-C2)	8	Corrente máxima negativa (I-max-neg-C1)
18	8	Corrente máxima negativa (I-max-neg-C1)	7	Corrente máxima positiva (I-max-pos-C1)
19	36	Corrente máxima negativa (I-max-neg-C3)	19	Total de raios positivos (total-raios-pos-C2)
20	35	Corrente máxima positiva (I-max-pos-C3)	22	Corrente máxima negativa (I-max-neg-C2)
21	21	Corrente máxima positiva (I-max-pos-C2)	36	Corrente máxima negativa (I-max-neg-C3)
22	7	Corrente máxima positiva (I-max-pos-C1)	34	Total de raios negativos (total-raios-neg-C3)
23	1	Data do evento (data-evento)	21	Corrente máxima positiva (I-max-pos-C2)
24	33	Total de raios positivos (total-raios-pos-C3)	15	Tempo pico a zero médio (tp0-med-C1)
25	32	Total de raios (total-raios-C3)	35	Corrente máxima positiva (I-max-pos-C3)
26	16	Tempo pico a zero mediana (tp0-mediana-C1)	26	Corrente mediana (I-mediana-C2)
27	34	Total de raios negativos (total-raios-neg-C3)	1	Data do evento (data-evento)
28	30	Tempo pico a zero mediana (tp0-mediana-C2)	30	Tempo pico a zero mediana (tp0-mediana-C2)
29	26	Corrente mediana (I-mediana-C2)	44	Tempo pico a zero mediana (tp0-mediana-C3)
30	44	Tempo pico a zero mediana (tp0-mediana-C3)	16	Tempo pico a zero mediana (tp0-mediana-C1)

Nesse estudo foi proposto um método para identificar as variáveis de entrada que deveriam ser excluídas do processamento do sistema, a fim de obter um melhor desempenho da rede e sua capacidade de generalização. Para essa seleção, em uma primeira análise foram considerados os três primeiros componentes principais. Em seguida, foram realizados os testes considerando os 8 primeiros componentes principais e por último, foram selecionados os 30 primeiros componentes principais.

5.4.3.1. Análise dos 3 Componentes Principais

Em uma primeira análise, foi realizada a ordenação dos *loadings* para identificar as variáveis mais relevantes, considerando os 3 componentes principais. Através da Figura 5.18 é possível observar uma grande relevância das variáveis correntes totais positiva e negativa. A análise dos 3 componentes principais definiu um conjunto de 10 variáveis de entrada, conforme apresentado na Tabela 5.11.

Tabela 5.11 - Indicação das variáveis de entrada mais relevantes - 3 CP

<i>Rank #</i>	<i>Entrada #</i>	Variável de Entrada	<i>Loading Ordenado</i>	<i>Loading Acumulado</i>
1	9	Corrente total positiva - I-total-pos-C1	0,99893117	26,07%
2	10	Corrente total negativa - I-total-neg-C1	0,99042782	51,92%
3	24	Corrente total positiva - I-total-neg-C2	0,85378905	74,20%
4	23	Corrente total negativa - I-total-pos-C2	0,51018008	87,52%
5	2	Número de raios - num-raios	0,10908806	90,37%
6	18	Total de raios C2 - total-raios-C2	0,08570127	92,60%
7	4	Total de raios C1- total-raios-C1	0,06677042	94,34%
8	6	Total de raios negativos C1 - total-raios-neg-C1	0,05261075	95,72%
9	20	Total de raios negativos C2 - total-raios-neg-C2	0,05012842	97,03%
10	19	Total de raios positivos C2 - total-raios-pos-C2	0,03566208	97,96%

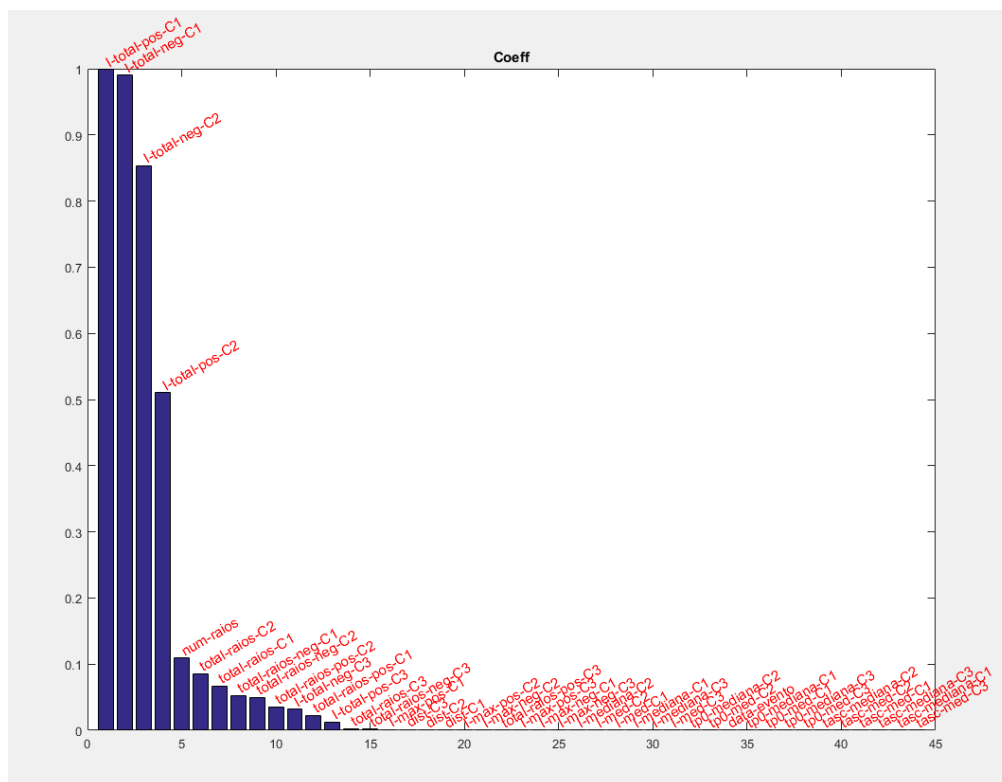


Figura 5.18 - Análise das variáveis de entrada - 3 Componentes Principais

5.4.3.2. Análise dos 8 Componentes Principais

Seguindo a análise da seleção de variáveis considerando as 3 primeiras componentes principais, foi observado os resultados considerando os 8 componentes principais. Com a realização da ordenação dos *loadings* para identificar as variáveis mais relevantes. Na Figura 5.19 são apresentadas as variáveis de entrada mais representativas, sendo identificadas as Correntes totais positiva e negativa, assim como na análise considerando as 3 componentes principais, sendo evidenciada também relevância ao número de raios do grupo. A análise dos 8 componentes principais definiu um conjunto de 16 variáveis de entrada, conforme apresentado na Tabela 5.12.

Tabela 5.12 - Indicação das variáveis de entrada mais relevantes - 8 CP

Rank #	Entrada #	Variável de Entrada	Loading Ordenado	Loading Acumulado
1	9	Corrente total positiva C1 - I-total-pos-C1	0,99997928	10,14%
2	23	Corrente total positiva C2 - I-total-pos-C2	0,99914596	20,26%

3	10	Corrente total negativa C1 - I-total-neg-C1	0,99900742	30,39%
4	37	Corrente total positiva C3 I-total-pos-C3	0,99897468	40,52%
5	24	Corrente total negativa C2 - total-neg-C2	0,99746746	50,63%
6	38	Corrente total negativa C3 -Itotal-neg-C3	0,99119196	60,68%
7	2	Número de raios - num-raios	0,83247507	69,12%
8	5	Total de raios positivos - total-raios-pos-C1	0,69913561	76,20%
9	4	Total de raios C1 - total-raios-C1	0,66074842	82,90%
10	18	Total de raios C2 - total-raios-C2	0,43054357	87,26%
11	6	Total de raios negativos C1 - total-raios-neg-C1	0,29002026	90,20%
12	19	Total de raios positivos - total-raios-pos-C2	0,27051147	92,95%
13	20	Total de raios negativos C2 - total-raios-neg-C2	0,17861963	94,76%
14	17	Distância do grupo C2 - dist-C2	0,09841954	95,76%
15	32	Total de raios C2 - total-raios-C3	0,08954134	96,66%
16	34	Total de raios negativos C3 - total-raios-neg-C3	0,07584555	97,43%

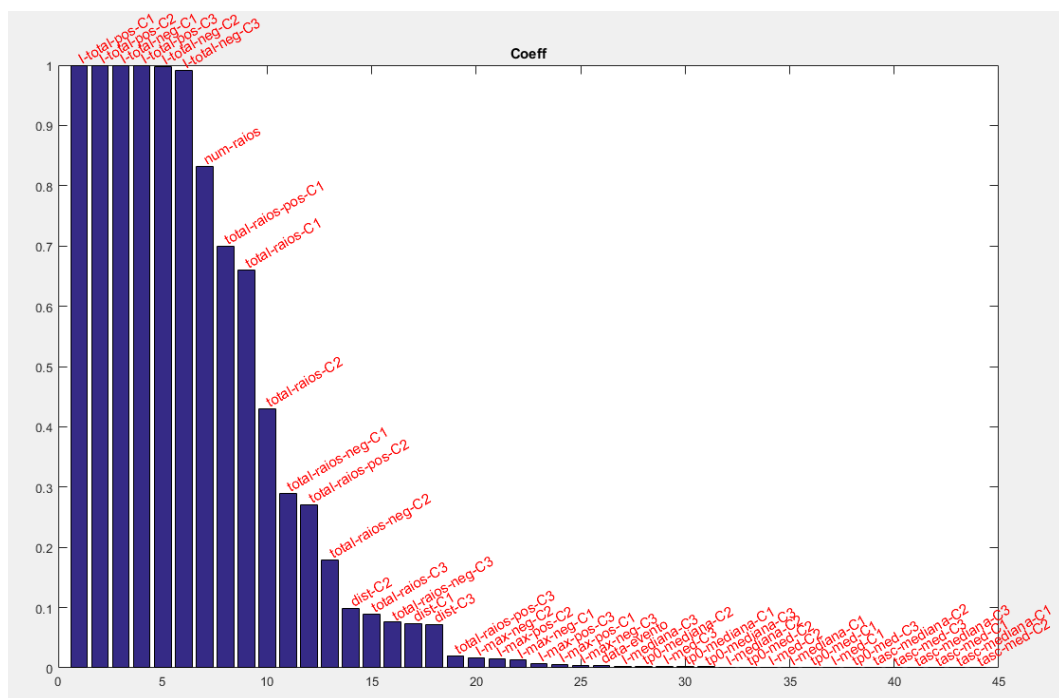


Figura 5.19 - Análise das variáveis de entrada - 8 Componentes Principais

5.4.3.3. Análise dos 30 Componentes Principais

No estudo para a identificação das variáveis mais relevantes considerando os 30 componentes principais, através da ordenação dos *loadings* foram obtidos os resultados apresentados na Tabela 20. Através da análise da Figura 5.20, é possível observar que se mantém como variáveis relevantes as correntes totais positivas e negativas, sendo identificado também o número de raios do grupo, além de passar a ter significância nessa análise a distância do grupo à linha de transmissão e total de raios positivos e negativos do grupo. A análise dos 30 componentes principais definiu um conjunto de 38 variáveis de entrada, conforme apresentado na Tabela 5.13.

Tabela 5.13 - Indicação das variáveis de entrada mais relevantes - 30**CP**

Rank #	Entrada #	Variável de Entrada	Loading Ordenado	Loading Acumulado
1	10	Corrente total positiva C1 - I-total-pos-C1	0,99999999	2,96%
2	9	Corrente total negativa C1 - I-total-neg-C1	0,99999978	5,92%
3	24	Corrente total positiva C2 - I-total-pos-C2	0,99999964	8,88%
4	23	Corrente total negativa C2 - I-total-neg-C2	0,99999898	11,84%
5	38	Corrente total negativa C3I-total-neg-C3'	0,9999942	14,80%
6	37	Número de raios - num-raios	0,99998716	17,75%
7	2	Corrente total positiva C3 - I-total-pos-C3	0,99998622	20,71%
8	5	Distância do grupo C3 - dist-C3	0,9998961	23,67%
9	18	Total de raios negativos C1 - total-raios-neg-C1	0,99963103	26,63%
10	6	Total de raios C1 - total-raios-C1	0,99956852	29,59%
11	4	Total de raios positivos C1 - total-raios-pos-C1	0,99946003	32,54%
12	32	Distância do grupo C1 - dist-C1	0,99913513	35,50%
13	20	Distância do grupo C2 - dist-C2	0,99908561	38,46%
14	31	Total de raios negativos C2 - total-raios-neg-C2	0,99784792	41,41%
15	17	Total de raios C2 - total-raios-C2	0,99689222	44,36%
16	3	Total de raios positivos C2 - total-raios-pos-C2	0,9967639	47,31%
17	8	Corrente máxima negativa C2 - I-max-neg-C2	0,99502941	50,25%
18	7	Corrente máxima negativa C1I-max-neg-C1	0,99044966	53,18%
19	19	Corrente máxima negativa C3 - I-max-neg-C3	0,98777937	56,11%
20	22	Corrente máxima positiva C3 - I-max-pos-C3	0,98522955	59,02%
21	36	Corrente máxima positiva C2 - I-max-pos-C2	0,98217236	61,93%
22	34	Corrente máxima positiva C1 I - max-pos-C1'	0,96916923	64,80%
23	21	Data do evento - data-evento	0,91903081	67,52%
24	15	Total de raios positivos C3 - total-raios-pos-C3	0,85320751	70,04%
25	35	Total de raios C3 - total-raios-C3	0,83798936	72,52%
26	26	Tempo de pico a zero mediana C1- tp0-mediana-C1	0,82833428	74,97%
27	1	Total de raios negativos C3 - total-raios-neg-C3	0,82390942	77,41%
28	30	Tempo de pico a zero mediana C2 - 'tp0-mediana-C2	0,82104621	79,84%
29	44	Corrente mediana C2 - I-mediana-C2	0,7110672	81,94%
30	16	Tempo de pico a zero mediana C3 - tp0-mediana-C3	0,70139652	84,02%
31	10	Corrente média C2 - I-med-C2	0,67866384	86,03%
32	9	Corrente mediana C1 - I-mediana-C1	6,70E-01	88,01%
33	24	Corrente média C3 - I-med-C3	6,68E-01	89,98%
34	23	Tempo de pico a zero médio C2 - tp0-med-C2	6,59E-01	91,93%
35	38	Corrente mediana C3 - I-mediana-C3	6,39E-01	93,82%
36	37	Tempo de pico a zero médio C3 - tp0-med-C3	5,26E-01	95,38%
37	2	Tempo de pico a zero médio C1 tp0-med-C1	4,40E-01	96,68%
38	5	Corrente média C1 I-med-C1	3,30E-01	97,66%

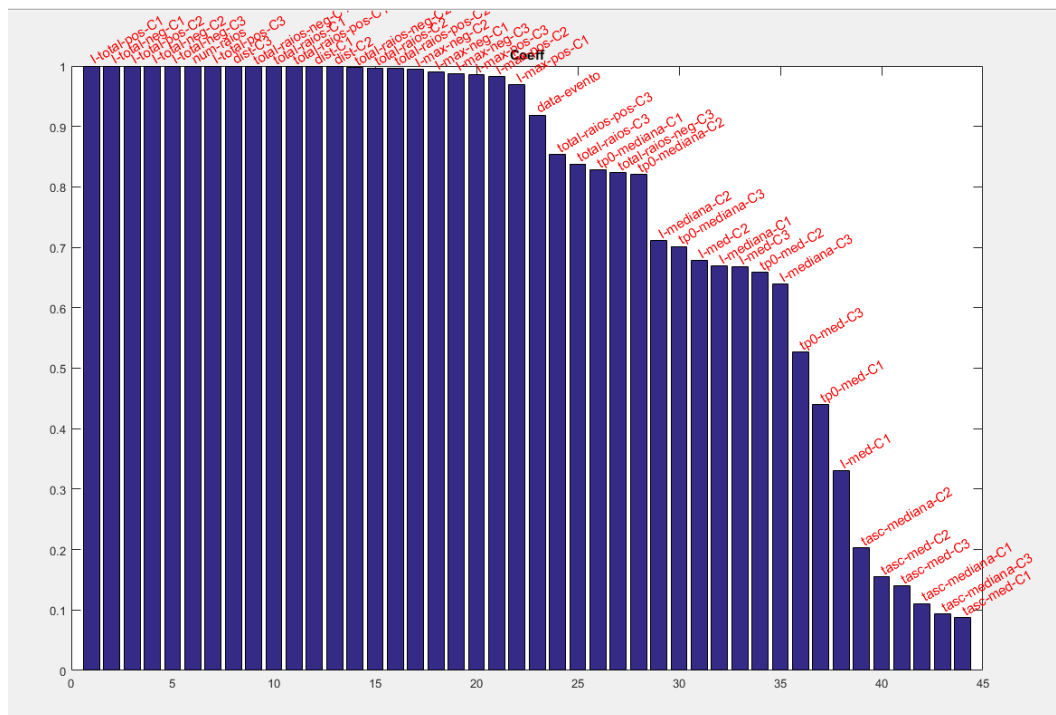


Figura 5.20 - Análise das variáveis de entrada - 30 Componentes Principais

5.4.4.

Definição da Camada Intermediária da Rede Neural MLP

O modelo da Rede Neural MLP não deve ser extremamente rígido, prejudicando o ajuste da modelagem dos dados e sua capacidade de aproximação, nem extremamente flexível, permitindo modelar também os ruídos naturalmente presente nos dados. Dessa forma, a seleção adequada do ajuste dos parâmetros para o treinamento da Rede Neural MLP é essencial para uma boa capacidade de generalização. O número de neurônios na camada intermediária é uma importante questão no projeto de uma arquitetura de uma Rede Neural MLP. Uma grande quantidade de unidades pode levar a rede a uma diminuição da capacidade de generalização (*overfitting*), impedindo a extração das características gerais. Por outro lado, um número muito pequeno de neurônios pode tornar a estrutura insuficiente para manipular as entradas, dificultando sobremaneira a convergência dos algoritmos de treinamento (*underfitting*).

Dessa forma, nesse estudo foram avaliadas diferentes configurações de número de neurônios na camada intermediária, buscando aquela que resulta em um menor erro médio sobre o conjunto de teste, definido após a execução de 50 instâncias no treinamento da Rede Neural MLP. Assim, foram executados o

treinamento da rede considerando 4, 6, 8, 10 e 12 neurônios na camada intermediária para a determinação do número de unidades que forneça o menor erro médio no conjunto de teste.

5.4.5. As Variáveis de Saída da Rede Neural MLP

A Rede Neural MLP nesse estudo possui duas classes, definidas como variáveis de saída dessa rede. Uma classe representa padrões de agrupamentos de descargas atmosféricas responsáveis por desligamento de linhas de transmissão (y_1) e a outra classe representa os agrupamentos de descargas atmosféricas não associados a desligamento de linhas de transmissão (y_2). São executadas 50 interações de treinamento, validação e teste, sendo que para cada uma delas, a rede neural calcula o valor das variáveis que representam as classes, referente a cada uma das 62 amostras. O valor de cada uma das variáveis de saída, compreendido entre 0 e 1, representa a capacidade classificatória da rede, permitindo entender padrões de descargas atmosféricas que de fato possam causar perturbações nas linhas de transmissão.

A Figura 5.21 representa a arquitetura da Rede Neural MLP utilizada no modelo RDLR.

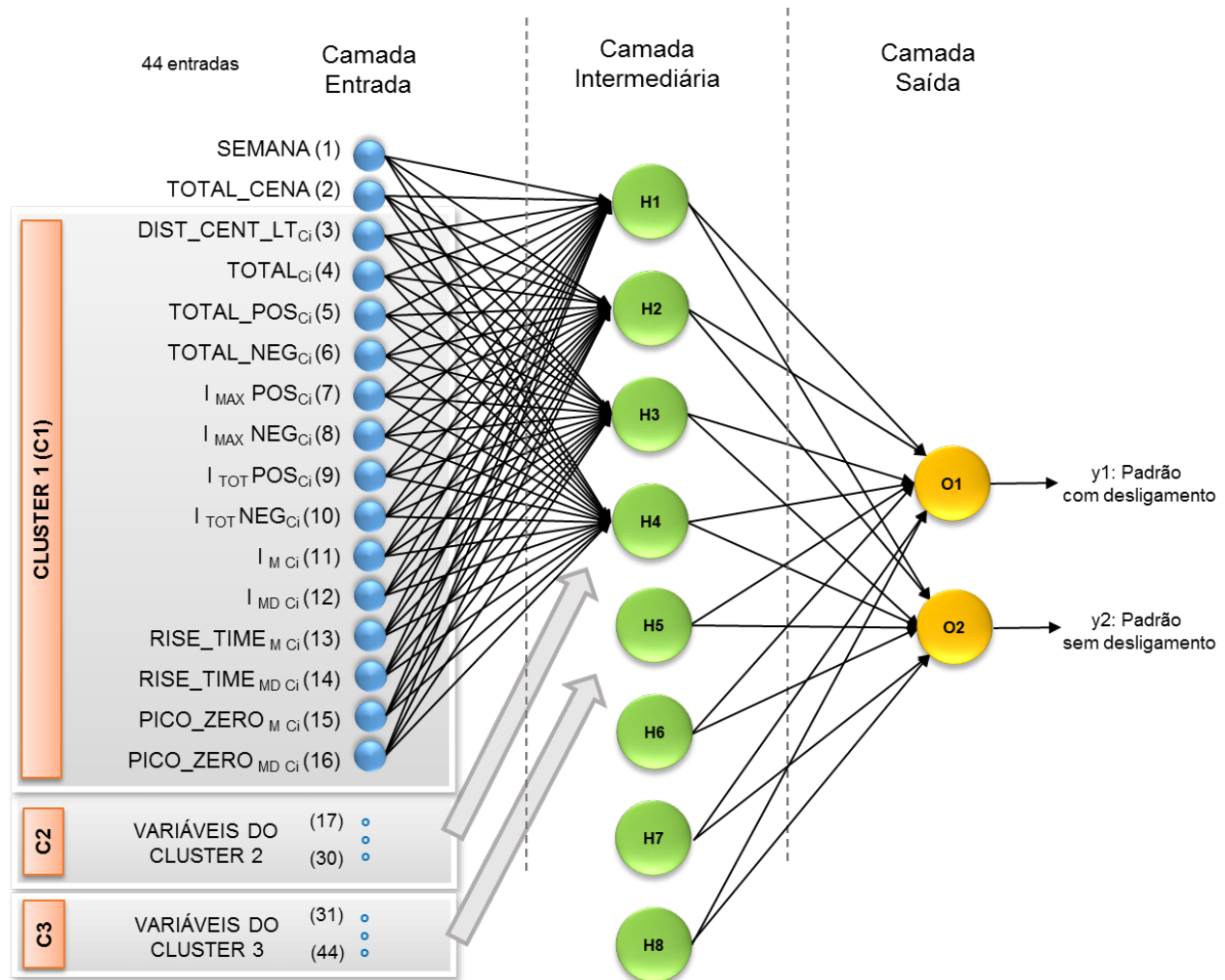


Figura 5.21 - A Rede Neural MLP do Modelo RDLR

5.4.6. A Determinação do Risco no Modelo RDLR

A partir dos valores das variáveis de saída da rede neural MLP, foi definido no modelo RDLR um método para determinar o risco do desligamento de linhas de transmissão por descargas atmosféricas. Conforme descrito no Capítulo 4, o método de determinação do risco consiste em um primeiro momento na comparação entre as duas variáveis de saída (y_1 e y_2), ou seja, nos resultados do desempenho das duas classes descritas no item anterior. A classe que produziu o maior valor absoluto é escolhida, e a partir dela o cálculo do risco é produzido.

Para as 62 amostras de descargas, 31 que causaram o desligamento e 31 que não causaram desligamento da linha de transmissão LT 440 kV TAQUARUÇU / JUPIA, o risco foi definido conforme condição:

Se $y_1 > y_2$, então

$$RDLR(\%) = y_1 \times 100\% \quad (10)$$

senão

$$RDLR(\%) = (1 - y_2) \times 100\% \quad (11)$$

onde,

RDLR - risco do desligamento da linha de transmissão

y_1 – variável de saída do padrão de desligamento da LT (classe 1)

y_2 – variável de saída do padrão não associado a desligamento da LT (classe 2)

O Modelo RDLR definido nesse estudo é representado através da Figura 5.22 e a análise dos resultados obtidos será apresentada no item a seguir.

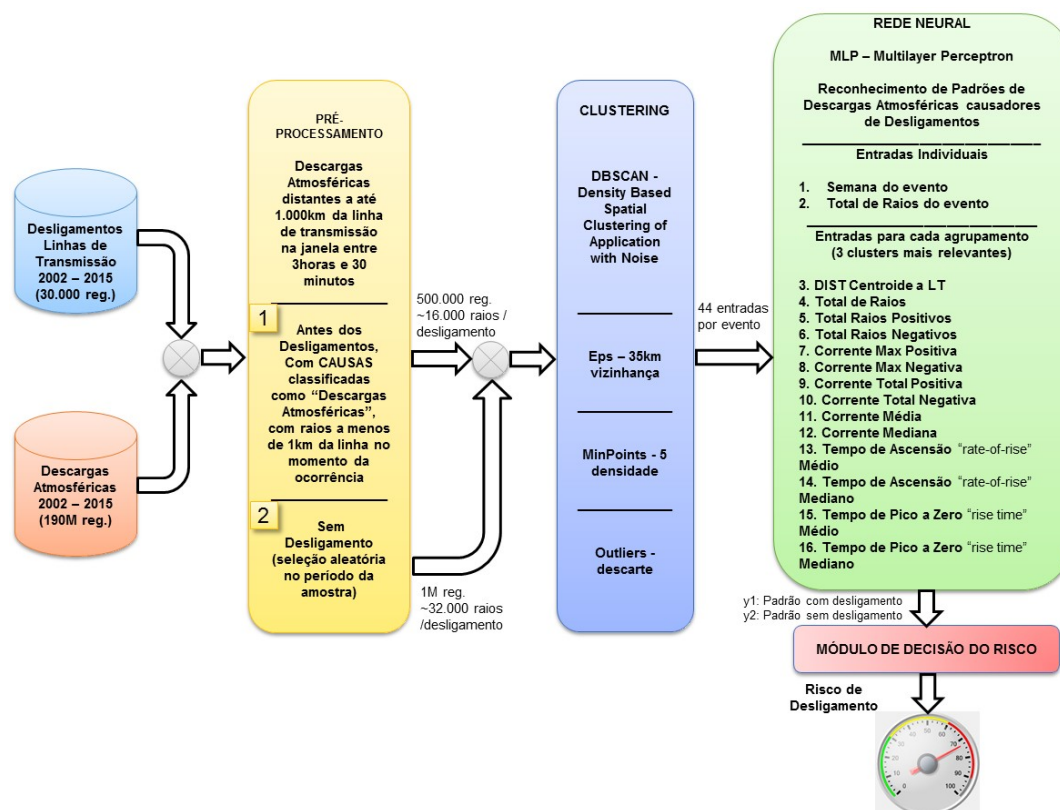


Figura 5.22 - Modelo RDLR para determinação do risco de desligamentos de linhas de transmissão por descargas atmosféricas.

5.5. Análise dos Resultados

Esse item trata da análise dos resultados obtidos pelo Modelo RDLR. Em uma primeira análise, foram observados os resultados comparativos relacionados às variáveis de entrada e sua seleção através da Análise dos Componentes Principais (PCA). No que tange as análises através da seleção de variáveis de entrada por PCA, serão também observados os resultados do método proposto nesse estudo, tratado no item 5.4.3 e o método PCA modificado. Assim, foram comparados os resultados utilizando as 44 variáveis de entrada, citadas no item 5.4.1 e as seleções dessas variáveis, através dos 3 componentes principais (10 variáveis de entrada), 8 componentes principais (16 variáveis de entrada), 30 componentes principais (38 variáveis de entrada) e 30 componentes principais do método PCA modificado (30 variáveis de entrada).

Os testes avaliaram também o número ideal de neurônios na camada intermediária, utilizando 4, 6, 8, 10 e 12 elementos, para as 62 amostras de descargas atmosféricas descritas nesse trabalho e 50 instâncias de processamento da Rede Neural MLP.

A Figura 5.23 apresenta os resultados das matrizes de confusão sobre o conjunto de teste, para as 25 configurações da Rede Neural MLP testadas nesse estudo, apresentando também a melhor instância dentre as 50 testadas para cada configuração. Dessas 25 configurações, 18 delas apresentaram 100% de acerto na classificação dos padrões de descargas atmosféricas responsáveis por desligamento da linha de transmissão e padrões não associados ao desligamento desse equipamento (classes 1 e 2), representando 72% das arquiteturas analisadas.

Com relação ao percentual de acertos totais por classe, das 25 arquiteturas estudadas, 17 obtiveram 100% de acerto para a classe 1 (padrões com desligamento) e 20 obtiveram 100% de acerto para a classe 2 (padrões sem desligamento).

Na arquitetura de 10 variáveis de entrada, apenas com 8 neurônios na camada intermediária houve 100% de acerto durante o processo classificatório. É importante notar que, para os demais testes com 4, 6, 10 e 12 elementos na camada intermediária, foram obtidos 88,9% de acerto nos resultados da rede

neural, representando apenas 1 das 9 amostras do conjunto de teste. Para os testes com 16 neurônios na camada de entrada, apenas com 4 e 8 neurônios na camada intermediária foram obtidos 100% de acerto nos resultados da rede neural, sendo observado 88,9% para os testes com 6, 10 e 12 neurônios na camada intermediária. Para as configurações da rede neural com 30, 38 e 44 variáveis de entrada, foram observados 100% de acerto da rede neural para todas as quantidades de neurônios na camada intermediária avaliadas. É possível observar que, conforme um número maior de variáveis de entrada, a rede neural apresentou melhores resultados.

RN10_4 Instância 48

RN10_6 Instância 24

RN10_8 Instância 46

RN10_10 Instância 36

RN10_12 Instância 32

RN16_4 Instância 46

RN16_6 Instância 37

RN16_8 Instância 20

RN16_10 Instância 16

RN16_12 Instância 20

RN30_4 Instância 16

RN30_6 Instância 11

RN30_8 Instância 50

RN30_10 Instância 16

RN30_12 Instância 34

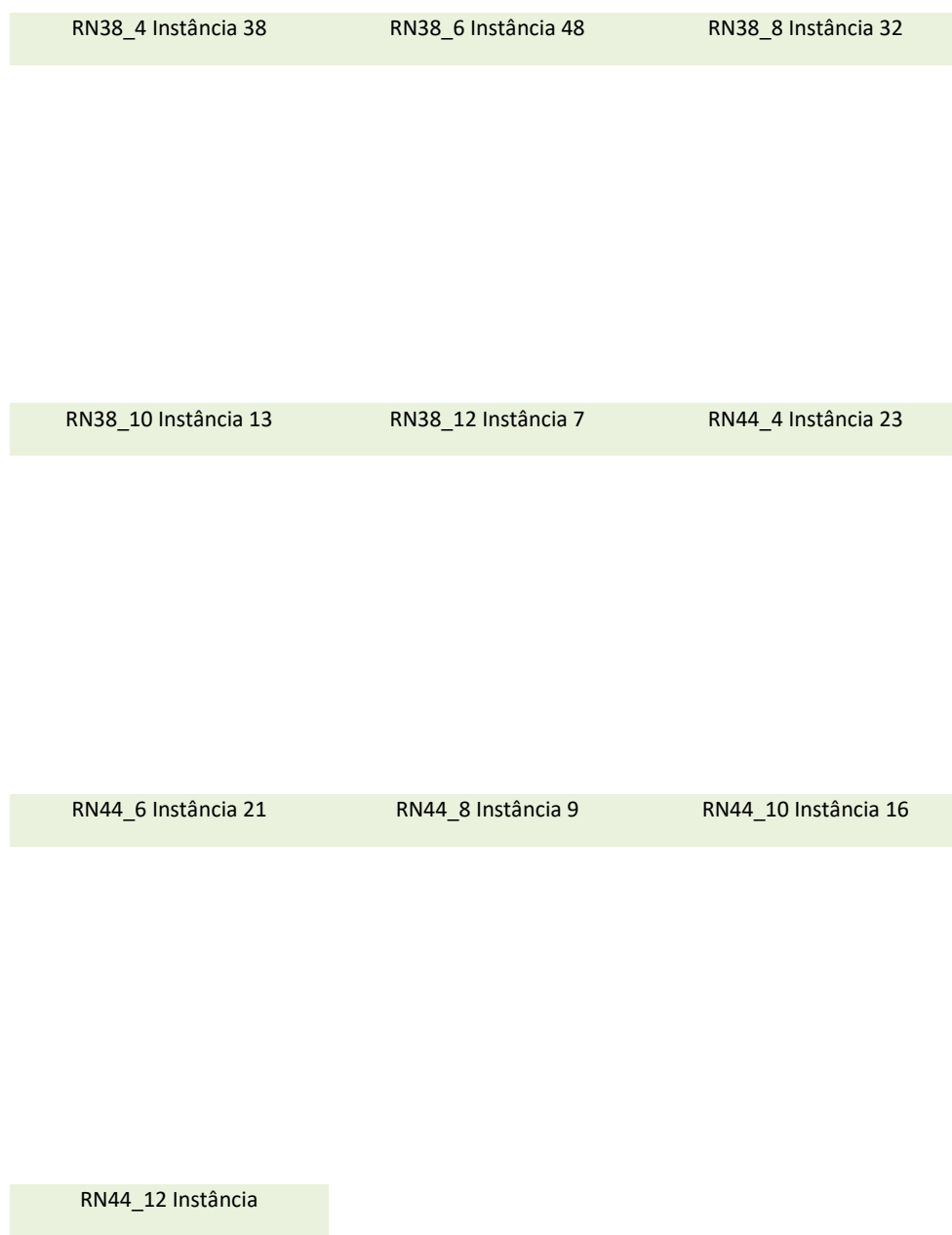


Figura 5.23 - Matrizes de Confusão sobre o conjunto de teste das 25 configurações de Rede Neural MLP e melhor instância

Após a avaliação dos resultados do desempenho classificatório da Rede Neural MLP, serão apresentados os resultados do Modelo RDLR.

A Tabela 5.14 apresenta o erro médio percentual do Modelo RDLR para as 50 instâncias sobre o conjunto de teste. É possível observar que o menor erro médio percentual foi obtido com 44 variáveis de entrada, sem a seleção de variáveis por PCA. Além dessa observação, o número de neurônios na camada intermediária que indicou o menor erro foi observado com 8 elementos, exceto nos testes realizados com 30 variáveis de entrada, através do PCA modificado, que apresentou o menor erro médio, com 6 neurônios na camada intermediária. A Figura 5.24 apresenta os resultados obtidos através de gráfico, observando o comportamento dos erros médios das 50 instâncias.

Tabela 5.14 - Erro médio das 50 instâncias do conjunto de teste

Erro Médio (%) - conjunto de teste					
Neurônios na camada intermediária	Número de Variáveis de Entrada				
	10 (PCA3) (%)	16 (PCA8) (%)	30 (PCA M) (%)	38 (PCA30) (%)	44 (%)
4	24,46	24,02	6,16	6,60	6,71
6	24,22	16,79	3,56	4,86	4,22
8	16,75	13,46	6,09	3,81	0,16
10	21,30	14,07	6,66	4,35	3,74
12	25,63	14,54	6,91	5,14	5,18

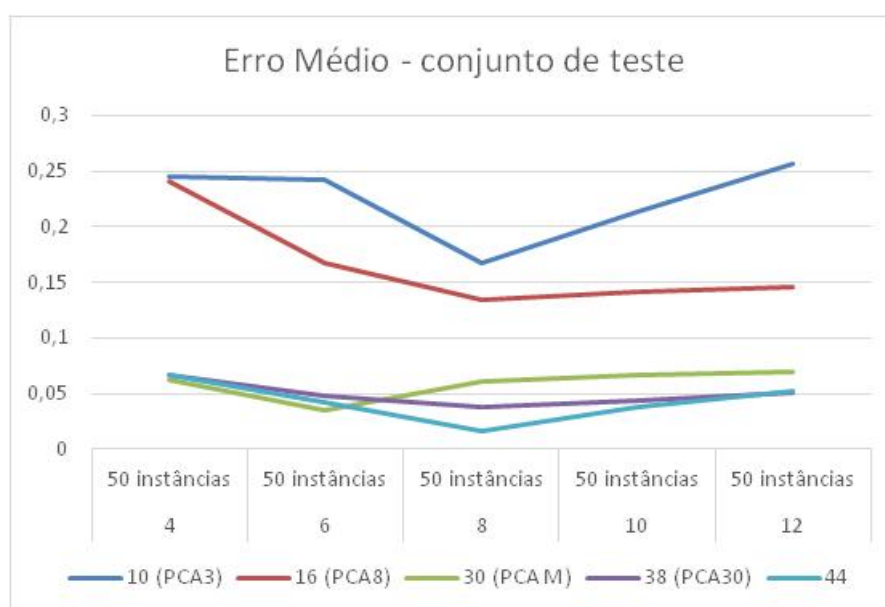


Figura 5.24 - Comportamento do erro médio das 50 instâncias sobre o conjunto de teste

Foram observados os resultados sobre o conjunto com todas as 62 amostras, obtendo o menor erro médio percentual das 50 instâncias também com 44 variáveis de entrada. Em todos os testes realizados, o número de neurônios com menor erro foi com 8 elementos, conforme Tabela 5.15. A Figura 5.25 apresenta os resultados obtidos através de gráfico, observando o comportamento dos erros médios percentuais das 50 instâncias sobre o conjunto com as 62 amostras.

Tabela 5.15 - Erro médio das 50 instâncias do conjunto total

Erro Médio (%) - conjunto total					
Neurônios na camada intermediária	Número de Variáveis de Entrada				
	10 (PCA3) (%)	16 (PCA8) (%)	30 (PCA M) (%)	38 (PCA30) (%)	44 (%)
4	24,35	23,58	5,19	4,65	4,66
6	24,10	14,09	4,65	4,07	2,18
8	15,55	12,32	4,32	2,73	0,77
10	18,08	13,05	4,52	2,86	1,92
12	22,38	14,53	5,96	3,80	4,78

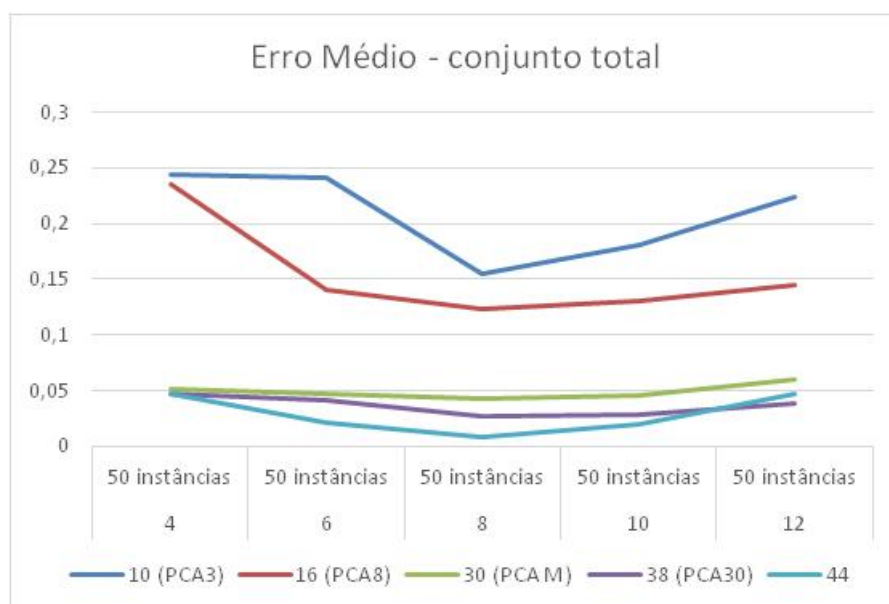


Figura 5.25 - Comportamento do erro médio das 50 instâncias sobre o conjunto total

Foi realizada uma análise com maiores detalhes sobre a instância que produziu melhores resultados sobre o conjunto de teste, através dos menores erros médios percentuais observados. Essa análise permite observar a instância em que o Modelo RDLR obteve o melhor desempenho, consequentemente os melhores parâmetros e ajustes da Rede Neural MLP. A Tabela 5.16 apresenta essas informações, mantendo as 44 variáveis de entrada com o menor erro médio percentual. O número de neurônios na camada intermediária que apresentou menor erro médio foi com 8 elementos, exceto nos testes realizados com 10 variáveis de entrada através de 3 componentes principais e com 30 variáveis de entrada selecionadas através do PCA Modificado. A Figura 5.26 apresenta os resultados obtidos através de gráfico, observando o comportamento dos erros médios percentuais da instância com melhor desempenho.

Tabela 5.16 - Erro médio da melhor instância do conjunto de teste

Erro Médio (%) - conjunto de teste					
Neurônios na camada intermediária	Número de Variáveis de Entrada				
	10 (PCA3) (%)	16 (PCA8) (%)	30 (PCA M) (%)	38 (PCA30) (%)	44 (%)
4	2,46	1,17	4,82E-06	1,31E-03	4,05E-05
6	0,01	0,09	1,32E-05	4,40E-06	1,08E-05
8	0,15	7,47E-06	6,29E-06	8,93E-07	8,15E-07
10	0,19	0,19	0,01	1,26E-06	3,14E-06
12	1,33	0,28	0,29	2,53E-05	0,05

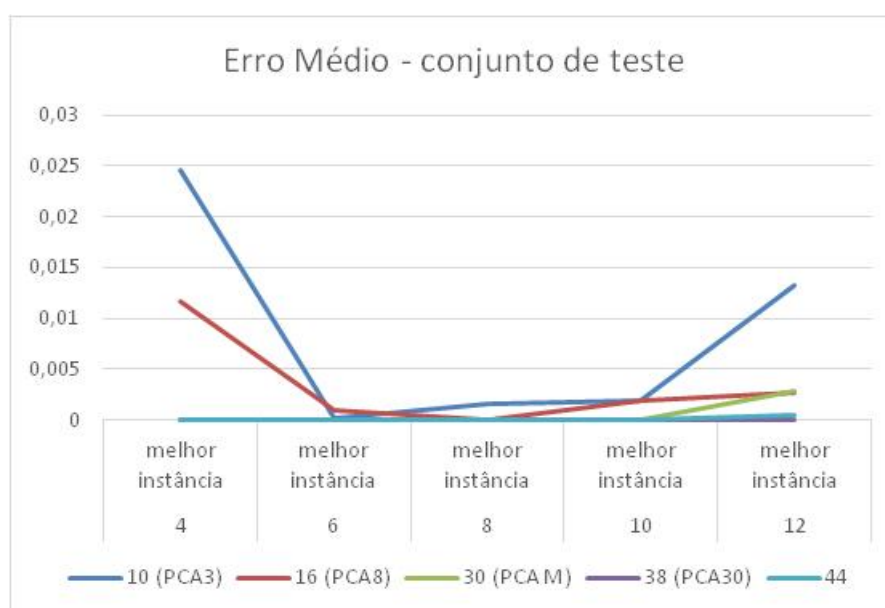


Figura 5.26 - Comportamento do erro médio da melhor instância sobre o conjunto de teste

Foram observados os resultados sobre o conjunto com todas as 62 amostras, obtendo o menor erro médio percentual da melhor instância, indicando o melhor resultado com 44 variáveis de entrada. Nos testes realizados, o número de neurônios com menor erro foi com 8 elementos, exceto nos testes realizados com 38 variáveis de entrada através de 30 componentes principais conforme Tabela 5.17. A Figura 5.27 apresenta os resultados obtidos através de gráfico, observando o comportamento dos erros médios percentuais da melhor instância sobre o conjunto com as 62 amostras.

Tabela 5.17 - Erro médio da melhor instância do conjunto total

Erro Médio (%) - conjunto total					
Neurônios na camada intermediária	Número de Variáveis de Entrada				
	10 (PCA3) (%)	16 (PCA8) (%)	30 (PCA M) (%)	38 (PCA30) (%)	44 (%)
4	21,16	8,16	3,26	5,63E-04	1,63
6	21,20	8,88	3,23	1,61	7,36E-04
8	11,15	8,06	0,01	0,33	2,88E-05
10	11,63	10,90	1,84	1,61	1,28E-03
12	20,43	12,09	3,81	3,23	1,96

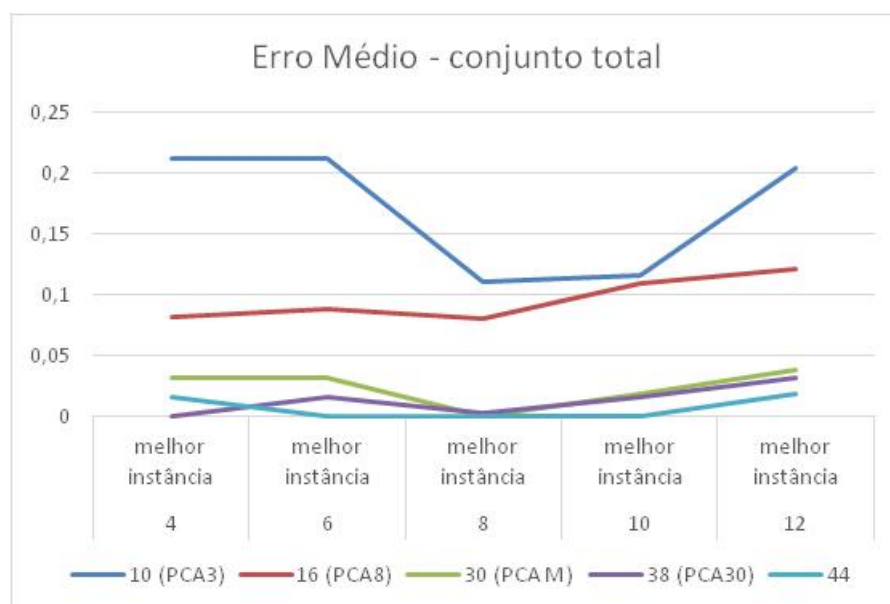


Figura 5.27 - Comportamento do erro médio da melhor instância sobre o conjunto total

A análise dos resultados permitiu concluir os melhores resultados do Modelo RDLR com 44 variáveis de entrada e 8 neurônios na camada intermediária. O menor erro médio percentual do conjunto de teste foi obtido na instância 18, conforme apresentado na Tabela 5.18. Os menores erros médios percentuais considerando todas as amostras foram obtidos das instâncias 44 a 50. O gráfico das Figuras 5.28 e 5.29 apresentam a redução do erro médio ao longo das instâncias. Esse fato é observado tanto para o conjunto de teste, como para o conjunto total com as 62 amostras. Desta forma, o desempenho do modelo estudado foi de 8,15E-07% na determinação do grau de risco da amostra no conjunto de teste, e uma taxa de acerto de 100% na classificação dos padrões com desligamento, obtidos na instância 18.

Tabela 5.18 - Erro médio dos conjuntos total e teste (44 variáveis de entrada e 8 neurônios na camada intermediária)

Instância	1	2	3	4	5	6	7	8	9	10
eMED teste	0,13191	0,028207	0,124707	0,118786	3,39E-02	0,119434	1,30E-01	9,75E-02	3,61E-07	2,66E-05
eMED total	0,057327	0,052823	0,052823	0,052823	0,051109	0,050425	0,050425	0,014157	0,003899	6,62E-06
Instância	11	12	13	14	15	16	17	18	19	20
eMED teste	4,81E-06	2,64E-05	2,03E-06	1,05E-05	3,01E-06	1,74E-05	1,66E-05	8,15E-09	3,47E-06	4,79E-08
eMED total	6,62E-06	6,62E-06	4,37E-06	4,37E-06	4,37E-06	3,84E-06	3,84E-06	3,84E-06	2,78E-06	2,78E-06
Instância	21	22	23	24	25	26	27	28	29	30
eMED teste	1,61E-05	2,44E-07	1,58E-07	7,13E-07	9,70E-08	1,92E-06	3,67E-06	1,56E-06	2,51E-07	1,75E-06
eMED total	2,85E-06	2,85E-06	1,35E-06	1,35E-06	1,35E-06	1,35E-06	1,35E-06	7,34E-07	6,42E-07	5,81E-07
Instância	31	32	33	34	35	36	37	38	39	40
eMED teste	5,90E-07	7,87E-08	2,94E-07	2,20E-07	4,95E-08	7,89E-07	5,33E-08	8,76E-07	2,80E-07	2,15E-07
eMED total	4,43E-07	4,43E-07	4,43E-07	4,43E-07	4,43E-07	4,43E-07	4,43E-07	4,03E-07	2,99E-07	2,99E-07
Instância	41	42	43	44	45	46	47	48	49	50
eMED teste	7,04E-07	6,74E-07	2,99E-07	1,47E-07	2,45E-07	6,66E-07	3,39E-07	1,00E-07	3,18E-07	2,71E-07
eMED total	2,99E-07	2,99E-07	2,99E-07	2,88E-07	2,88E-07	2,88E-07	2,88E-07	2,88E-07	2,88E-07	2,88E-07

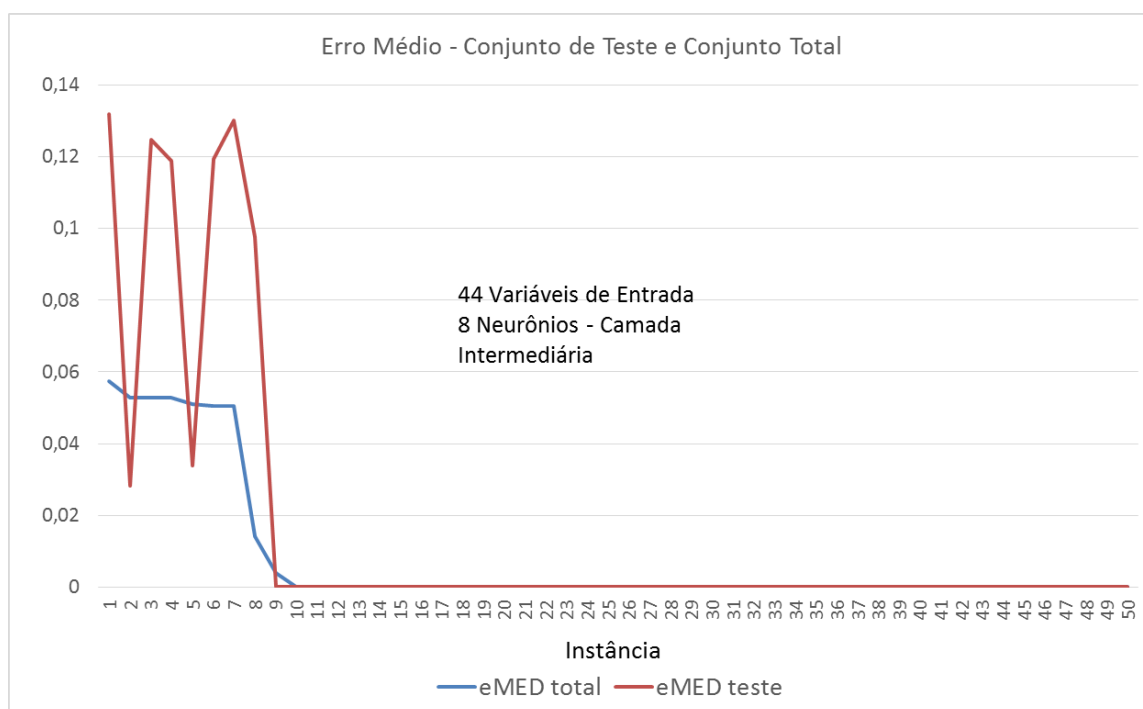


Figura 5.28 - Erro médio dos conjuntos de teste e total (44 variáveis de entrada e 8 neurônios na camada intermediária)

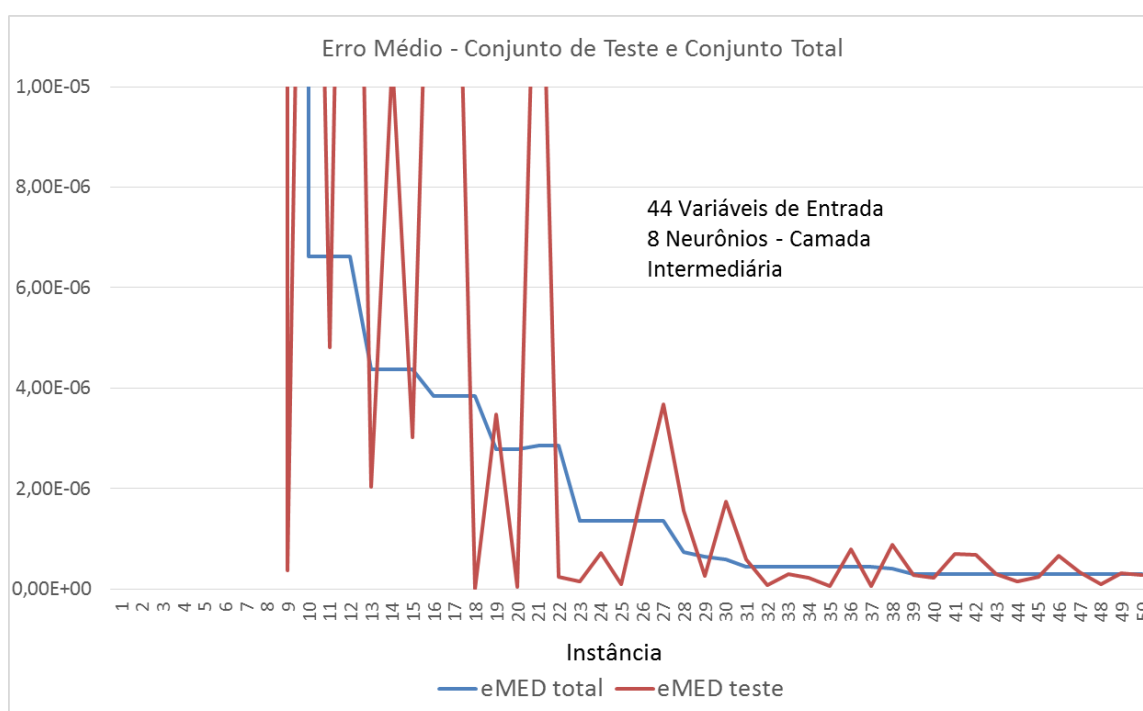


Figura 5.29 – Detalhe do erro médio dos conjuntos de teste e total (44 variáveis de entrada e 8 neurônios na camada intermediária)

6

Conclusões e Recomendações

Esta dissertação apresentou a proposição de um modelo para a determinação em tempo real dos riscos de desligamentos em linhas de transmissão, devido a descargas atmosféricas, baseado em Redes Neurais Artificiais, fornecendo informações antecipadas e possibilitando ações operativas para manter a segurança do sistema elétrico.

O estudo permitiu correlacionar os 29262 registros sobre desligamentos de linhas de transmissão, ocorridos entre os anos de 2002 a 2015 em todo território brasileiro, com os 190 milhões de registros de descargas atmosféricas obtidos através das redes de detecção, armazenados em base histórica desde 2001.

A avaliação positiva sobre a tendência de agrupamento dos dados de descargas atmosféricas através da Estatística de Hopkins comprovou sua forte característica de dados agrupáveis ($H \sim 1$). Reforçando os resultados sobre tendência de agrupamento através da Estatística de Hopkins, o algoritmo DBSCAN utilizado nesse estudo gerou agrupamentos de diferentes formas e tamanhos, eliminando os ruídos com um bom desempenho.

Durante a definição dos critérios de seleção de linhas de transmissão para estudo, observou-se que em regiões com grande ocorrência de desligamentos em linhas de transmissão por descargas atmosféricas no setor de 500kV, nas regiões centro-oeste e nordeste do país, ainda não possuem uma detecção eficiente das descargas, em função da baixa quantidade de sensores da rede nestas localidades.

O Modelo RDLR apresentou excelentes resultados nesse estudo, sendo obtido, na melhor instância da rede neural MLP, um erro médio de $8,15E-09$ na determinação do grau de risco da amostra no conjunto de teste, e uma taxa de acerto de 100% na classificação dos padrões com desligamento. A partir desses bons resultados obtidos com o Modelo RDLR, são recomendados os seguintes trabalhos futuros:

- A Rede Neural MLP do Modelo RDLR pode ser treinada, validada e testada com novos registros de desligamentos de linhas de transmissão por descargas atmosféricas após o período definido nesse estudo, a fim de uma contínua melhoria do modelo.
- Esse estudo foi produzido a partir de informações sobre as causas associadas aos desligamentos das linhas de transmissão classificadas como "Descargas Atmosféricas". Porém, outras causas que indiretamente possam ter associação a incidência de raios, tal como "Chuva/Temporal", "Vento Forte" e "Indeterminada", também ocorrem com bastante frequência. Portanto, visando o aumento da amostra de cenas de desligamento por descargas, deve ser verificada a incidência de raios em ocorrências dessas naturezas.
- Esse trabalho tratou das cenas com dados de descargas atmosféricas tipo nuvem-terra, por serem os eventos efetivamente responsáveis pelos desligamentos de linhas de transmissão. No entanto, a ocorrência de um grande volume de descargas nuvem-nuvem, frequentemente antecede a incidência de descargas nuvem-terra. Diante dessa característica, será importante avaliar em estudos futuros os padrões de descargas nuvem-nuvem para os cenários onde há desligamentos de linhas de transmissão e naqueles onde não há perturbações nas linhas de transmissão devido a raios.
- O estudo da seleção de variáveis nesse trabalho, através da Análise de Componentes Principais (PCA), mostrou resultados bastante interessantes, podendo haver um estudo mais aprofundado com a exploração de outros métodos existentes para produzir resultados comparativos.

7

Referências Bibliográficas

[1] Visacro, Silvério, “Descargas Atmosféricas: Uma Abordagem de Engenharia”, Editora Líber, 2005.

[2] ANEEL, “Resolução Normativa nº 270 de 26 de junho de 2007”, Agência Nacional de Energia Elétrica, 2007.

[3] Morales, Carlos A.; Sales, Francisco; Pinheiro, F.Geraldo; Câmara, Keyla S., “sistema de monitoramento de descargas atmosféricas para os estados do Ceará”, 2008.

[4] Cummins, K.L., M. Murphy, E. Bardo, W. Hiscox, R. Pyle, and A. Pifer, 1998a: A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network, J. Geophys. Res., 103, 9035-9044.

[5] Lima, W.F.A., Machado, L.A.T., Morales, C.A., Pinto Jr., O.: Estimativa de probabilidade de ocorrência de relâmpagos, XIV Congresso Brasileiro de Meteorologia: A Meteorologia a Serviço da Sociedade, Florianópolis – SC, 27.11 – 01.12.2006, CD-ROM.

[6] Ary D’Ajuz, “Transitórios Elétricos e Coordenação de Isolamento – aplicação em sistemas de potência de alta tensão”- EDUFF – 1987.

[7] UMAN, M.A., “Lightning”, Livro, Editora McGraw-Hill Book Company, 1984.

[8] Amilton Soares Júnior – “Modelagem de Linhas de Transmissão para Avaliação de Desempenho frente a Descargas Atmosféricas” – Tese de Doutorado, Universidade Federal de Minas Gerais – 2001.

[9] SALANAVE, L. Lightning and its spectrum. 1980 Ogawa, T. Lightning currents. In: Volland, H. (ed.). Handbook of atmospheric electrodynamics. Boca Raton: CRC Press, 1995. v. 1, Cap. 4, p. 93-136.

[10] NACCARATO, K.P.; SARAIVA, A.C.V.; SABA, M. M.F.; SCHUMANN, C. First performance analysis of BrasilDAT total lightning network In southeastern Brazil. In: GROUND 2012 INTERNATIONAL CONFERENCE ON GROUNDING AND EARTHING & 5TH INTERNATIONAL CONFERENCE ON LIGHTNING PHYSICS AND EFFECTS, 2012, Bonito - MS. Proceedings... Bonito, 2012.

[11] PLANO DE AÇÃO DO ONS 10.21 – SISTEMAS DE INFORMAÇÕES SOBRE DESCARGAS ATMOSFÉRICAS (GAT-CT-169/07) - RELATÓRIO COM EMBASAMENTO TÉCNICO

[12] KINDERMANN, Geraldo. Choque Elétrico. Porto Alegre: Editora Sagra DC Luzzato, 2. ed., 2000.

[13] POTIER, Guido de Camargo. Sobre os Raios. Porto Alegre, 2003. Disponível em Acesso em 19 dez. 2007.

[14] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 5419: Proteção de estruturas contra descargas Atmosféricas. 2005.

[15] IEEE Guide for Improving the Lightning Performance of Electric Power Overhead Distribution Lines – IEEE Std 1410 – 2004.

[16] BRASILDAT. Sistema Brasileiro de Detecção de Descargas Atmosféricas. 2017. Disponível em: <http://www.ccst.inpe.br/projetos/brasildat>

[17] RINDAT. Rede Integrada Nacional de Detecção de Descargas Atmosféricas. 2017. Disponível em: <http://simepar.br/rindat/internas/institucional.shtml>

[18] MIKI, M.; SHINDO, T.; RAKOV, V.A.; UMAN, M.A.; DIENDORFER, G.; MAIR, M.; HEIDLER, F.; ZISCHANK, W.; THOTTAPPILLIL, R. Characterization of current pulses superimposed on the continuous current in upward lightning initiated from tall objects and in rocket-triggered lightning. In: INTERNATIONAL 75

CONFERENCE ON LIGHTNING PROTECTION, 28., 2006, Kanazawa, Proceedings... Kanazawa: ICLP Centre, 2006.

[19] Kumar, V., Tan, P., and Steinbach, M. (2006). Cluster Analysis: Basic Concepts and Algorithms. In Introduction to Data Mining, chapter 8, pages 487–586.

[20] Andritsos P., Data Clustering Techniques Qualifying Oral Examination Paper, University of Toronto Department of Computer Science periklis@cs.toronto.edu March 11, 2002.

[21] Fung G, A Comprehensive Overview of Basic Clustering Algorithms, June 22, 2001.

[22] HAN, J.; KAMBER, M.; TUNG, A. K. H. Spatial clustering methods in data mining: a Survey. School of Computing Science, Simon Fraser University, Burnaby, BC Canada, v5a156, 2001.

[23] HAN, J.; FU, Y.; HUANG, Y.; CAI, Y.; CERCONE, N. DBLearn: a system prototype for knowledge discovery in relational databases. In Proceedings... on ACM SIGMOD International Conference on Management of Data, 516. ACM, 1994.

[24] KOHONEN, T. Self-organized formation of topologically correct feature maps. Biological Cybernetics, v.43, p.59-69, 1982.

[25] JAIN, A. K., DUBES, R. C. Algorithms for Clustering Data. Englewood Cliffs, Prentice-Hall, 1988.

[26] HAYKIN, S. Redes neurais: princípios e prática. 2.ed. Tradução de, Paulo Martins Engel. Porto Alegre: Bookman, 2001. 900 p., il. Obra originalmente publicada sob o título: Neural Networks: a comprehensive foundation, 2/E.

[27] JAIN, A. K., Murty, M.N., Flynn, P.J. Data Clustering: a review. ACM Comp. Surv., v.31, n.3, p. 264-323, 1999.

[28] Procedimentos de Rede do ONS, Submódulo 2.4 – Requisitos mínimos para linhas de transmissão aérea, Revisão 2.0.

[29] Lutgens, F. K.; Tarbuck, J. E. The Atmosphere. 2 ed. Englewood Cliffs: Prentice-Hall, 1982.

[30] CUMMING, J.A., WOOFF, D.A., 2007. Dimension reduction via principal variables. Computational Statistics & Data Analysis, v. 52, pp. 550 – 565.

[31] SILVA, A.P.D., 2001. Efficient Variable Screening for Multivariate Analysis. Journal of Multivariate Analysis, v. 76, pp. 35-62.

[32] LUZ, E. R. Predição de propriedades de gasolinas usando espectroscopia FTIR e regressão por mínimos quadrados parciais. 2003. Dissertação (Mestre em Química). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2003.

[33] JOLLIFFE I.T., 1972. Discarding Variables in a Principal Component Analysis. I: Artificial Data. Journal of the Royal Statistical Society. Series C (Applied Statistics), v. 2, n. 21, pp. 160-173.

[34] ENDSLEY, M. R., BOLTÉ, B., & JONES, D. G. (2003). Designing for situation awareness: An approach to user-centered design. New York: Taylor & Francis.

ANEXO I.

Aspectos Teóricos sobre Agrupamento

A. Agrupamento dos Dados

Agrupamento é a divisão de dados, com base na similaridade entre eles, em grupos disjuntos chamados “*clusters*”. Isso significa que dados em um mesmo grupo são mais similares do que dados pertencentes a grupos diferentes. O ato de agrupar os dados representa uma ferramenta importante no aprendizado e compreensão a respeito dos mesmos.

Pode ser definido também como um problema de aprendizado não-supervisionado, já que a estrutura dos dados e as propriedades que o tornam similares são desconhecidas. O objetivo do agrupamento é encontrar uma organização válida e conveniente dos dados, ao invés de separá-los em categorias como acontece no reconhecimento de padrões e na classificação de dados (Jain&Dubes, 1988).

Distinta do conceito de classificação, o agrupamento é uma técnica mais primitiva, na qual nenhuma suposição é feita a respeito dos grupos. Ao contrário da classificação, o agrupamento não possui classes predefinidas e exemplos de treinamento de classes rotuladas, e dessa forma, realiza uma forma de aprendizado não supervisionado.

O primeiro registro publicado sobre um método de agrupamento foi feito em 1948, com trabalho de SORENSEN (1948) sobre o Método Hierárquico de Ligação Completa. Desde então, mais de uma centena de algoritmos distintos de agrupamento já foram definidos. Qualquer método de agrupamento é definido por um algoritmo específico que determina como será feita a divisão dos casos nos grupos distintos e todos os métodos propostos são fundamentados na ideia de distância ou similaridade entre as observações, definindo a pertinência dos objetos a cada grupo, segundo o que cada elemento possui de similar em relação a outros pertencentes do grupo.

Os elementos que compõem um mesmo grupo devem apresentar alta similaridade e dissimilaridades de objetos de outros grupos. Ou seja, todo agrupamento é realizado com objetivo de maximizar a homogeneidade dentro de cada grupo e maximizar a heterogeneidade entre grupos.

A grande vantagem do uso das técnicas de agrupamento é que, ao agrupar dados similares, pode-se descrever de forma mais eficiente e eficaz as

características peculiares de cada um dos grupos identificados. Isso fornece um maior entendimento do conjunto de dados original, além de possibilitar o desenvolvimento de esquemas de classificação para novos dados e descobrir correlações interessantes entre os atributos dos dados que não seriam facilmente visualizadas sem o emprego de tais técnicas. Alternativamente, o agrupamento pode ser usado como uma etapa de pré-processamento para outros algoritmos, tais como caracterização e classificação, que trabalhariam nos grupos identificados.

Grupos de dados de diferentes formas, densidades e tamanhos estão presentes em diferentes conjuntos de dados. Um objeto não pode pertencer a mais de um grupo (grupos disjuntos) e cada grupo deve ser formado pelo menos por um objeto.

B. Aplicações

O agrupamento pode ser empregado quando o objetivo é reduzir o número de objetos, para um número de subgrupos característicos, levando as observações a serem consideradas como membros de um grupo e alinhadas segundo características gerais que rotulam distintamente este grupo. Outra utilidade do agrupamento é a identificação de relacionamentos entre as observações não identificadas em outras técnicas. Entretanto, o uso mais tradicional do agrupamento tem sido para propósitos exploratórios e formação de uma taxonomia, uma classificação de objetos com base empírica.

A popularização do uso e desenvolvimento de métodos de agrupamento ocorrem devido à grande quantidade de dados coletados nas diversas áreas do conhecimento e atividade que tornam a análise dos grupos um tópico altamente atrativo em várias pesquisas na mineração de dados.

Como exemplos de áreas interessadas no problema do agrupamento, podemos citar: mineração de dados, estatística, engenharia, aprendizado de máquina, medicina, marketing, administração e biologia. As aplicações para o agrupamento são utilizadas para cumprir pelo menos um dos seguintes objetivos principais:

- Identificação da estrutura subjacente: para obter informação sobre os dados, gerar hipóteses, detectar anomalias, e identificar características marcantes.
- Classificação natural: identificar o grau de semelhança entre as formas ou organismos (filogenética).
- Compressão: como um método para a organização dos dados e resumindo-o através de protótipos do grupo.

C. O Problema do Agrupamento

O termo “clusterização” vem do inglês to cluster, que significa “agrupar”, “aglomerar”. Assim, no presente contexto, se faz o ato de agrupar dados. O agrupamento de dados, segundo Cortês et al (2002) é uma técnica que visa detectar a existência de diferentes grupos dentro de um determinado conjunto de dados e, em caso da existência, determinar estes grupos.

Nos casos de conjunto de dados com informações visíveis relativas aos dados apresentados, aliados a uma pequena quantidade desses, é evidente a facilidade de identificar os indivíduos com características comuns. Porém, há situações em que os dados não contém informações tão evidentes ou conhecidas. Uma análise mais profunda seria necessária para conseguir separar os dados em grupos afins.

ANKERST et al. (1999) escrevem que existem três razões interconectadas para explicar porque a efetividade dos algoritmos de agrupamento é um problema. Primeiro, quase todos os algoritmos de agrupamento requerem valores para os parâmetros de entrada que são difíceis de determinar, especialmente para conjuntos de dados do mundo real contendo objetos com muitos atributos. Segundo, os algoritmos são muito sensíveis a estes valores de parâmetros, frequentemente produzindo partições muito diferentes do conjunto de dados mesmo para ajustes de parâmetros significativamente pouco diferentes. Terceiro, os conjuntos de dados reais de alta dimensão têm uma distribuição muito ampla que não pode ser revelada por um algoritmo de agrupamento usando somente um ajuste de parâmetro global.

D. Etapas do Processo de Agrupamento

Diversos fatores devem ser levados em consideração para assegurar a eficiência do agrupamento: coleta de dados, seleção de variáveis, medidas de similaridade, algoritmo de agrupamento, validação e análise dos resultados obtidos (Jain & Dubes, 1988), conforme Figura 1.

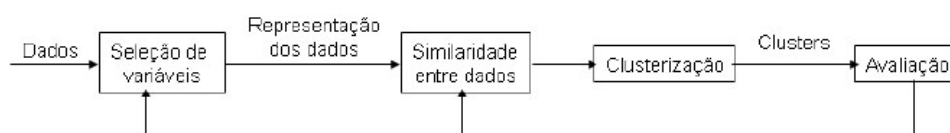


Figura 1 – Etapas do processo de agrupamento

A seguir são apresentadas as etapas que compõem o processo de agrupamento:

Pré-processamento e seleção de variáveis: Etapa que são identificadas as variáveis, ou atributos, mais relevantes do conjunto de dados inicial. Nessa etapa algumas variáveis podem ser eliminadas (por exemplo, quando todos os valores da variável são iguais) e os dados são coletados de modo que o algoritmo de agrupamento possa processá-los.

Medidas de similaridade: Para que a proximidade entre dois dados possa ser quantificada, é necessário adotar alguma medida de similaridade entre elas. Existem diversas maneiras de quantificar a similaridade, entre pares de dados e a escolha da medida de similaridade adequada é fundamental para o agrupamento dos dados.

Algoritmos de agrupamento: Nessa etapa define-se o modo de agrupamento dos dados, que pode ser realizado de diferentes maneiras. Os algoritmos de agrupamento são classificados de acordo com as diferentes técnicas que

empregam no agrupamento dos dados. O resultado dessa etapa é a divisão do conjunto de dados inicial em grupos.

Validação e análise dos resultados: Nessa etapa a qualidade dos grupos encontrados é avaliada, já que são desconhecidos inicialmente. Essa validação pode ser feita com base em índices estatísticos ou através de comparação com outros algoritmos. Além disso, a análise dos resultados pode levar à redefinição das medidas de similaridade, definidas previamente.

ANEXO II.**Avaliação da Tendência de Agrupamento dos Dados
de Descargas Atmosféricas**

E. Tendência de Agrupamento

Uma importante etapa no agrupamento é avaliar primeiramente a tendência dos dados formarem grupos, ou seja, se uma determinada amostra de dados possui agrupamentos significativos. Se um algoritmo apresenta grupos com boa coesão e boa separação pode-se concluir que os dados apresentam tendência de agrupamento. Portanto, uma avaliação prévia da tendência de agrupamento dos dados é uma etapa essencial, evitando aplicações inapropriadas dos algoritmos de agrupamento. Um algoritmo de agrupamento formará grupos se os dados apresentarem uma natureza puramente randômica.

É possível utilizar uma abordagem estatística para avaliação da tendência de agrupamento dos dados, como a equação de Hopkins [12]. Nesta técnica, deve-se gerar p registros uniformemente distribuídos pelo espaço e utilizar p registros da base real.

$$H = \frac{\sum_{i=1}^p w_i}{\sum_{i=1}^p u_i + \sum_{i=1}^p w_i} \quad (12)$$

Sendo u_i as distâncias da vizinhança mais próxima dos dados gerados artificialmente, e w_i dos dados reais. Se u_i e w_i forem semelhantes (significa que o agrupamento está próximo dos dados gerados artificialmente), H deve ser próximo de 0,5. Valores de H próximos de 0 indicam que os dados não são agrupáveis, e próximos de 1 que os dados possuem tendência a agrupamento. Essa equação mede a diferença entre os objetos do conjunto e objetos aleatórios, ou seja, mede a aleatoriedade dos objetos, conforme figuras abaixo.

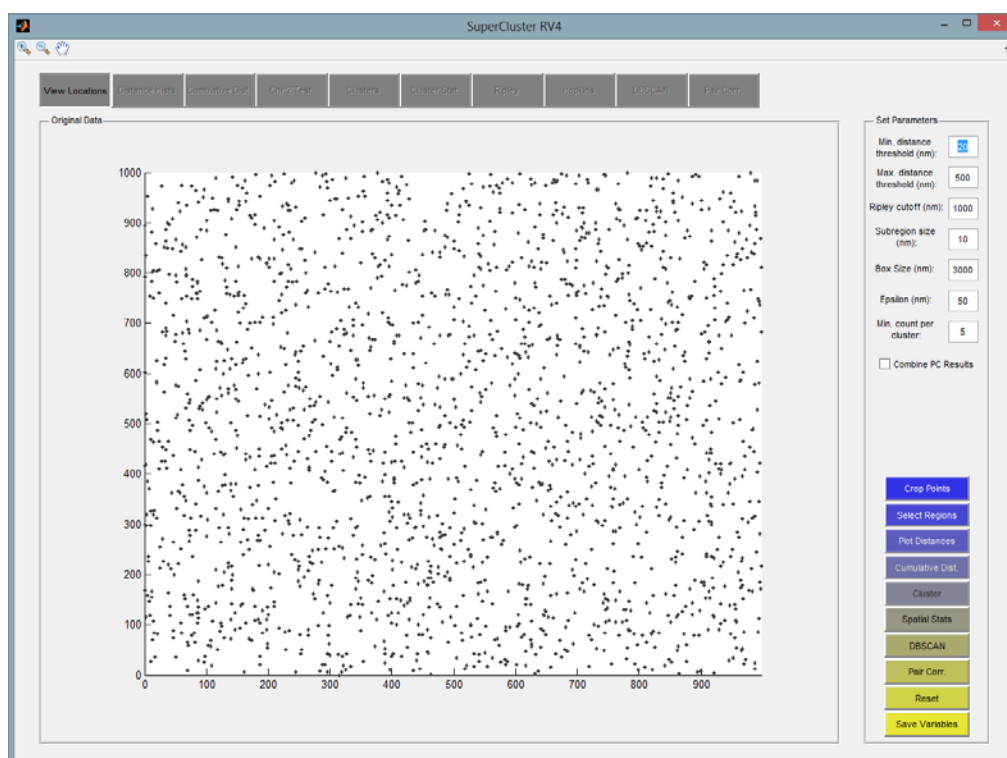


Figura 2 – Amostra randomicamente distribuída no espaço

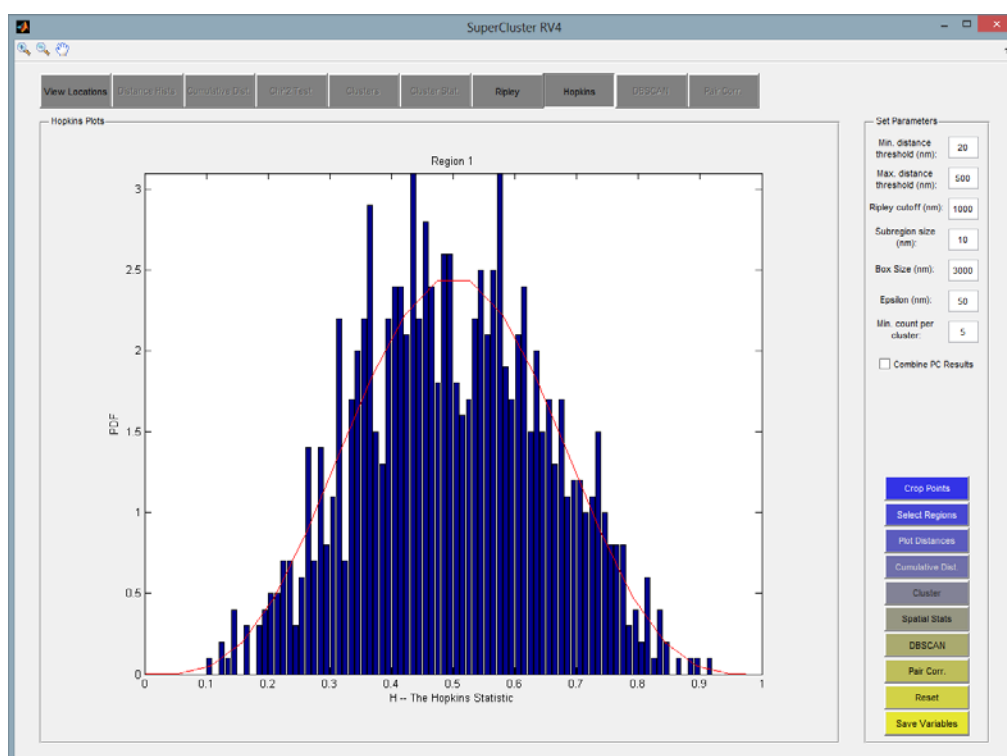


Figura 3 – Estatística de Hopkins para amostra randomicamente distribuída no espaço ($H \approx 0,5$)

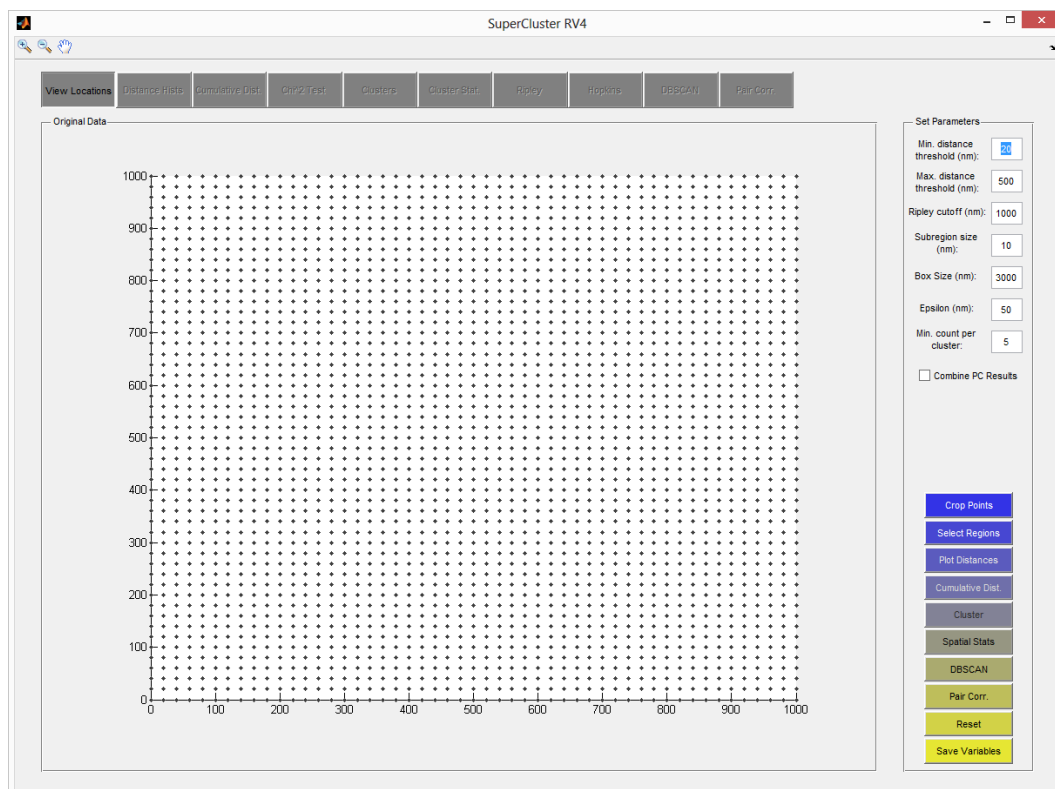


Figura 4 – Amostra uniformemente distribuída no espaço

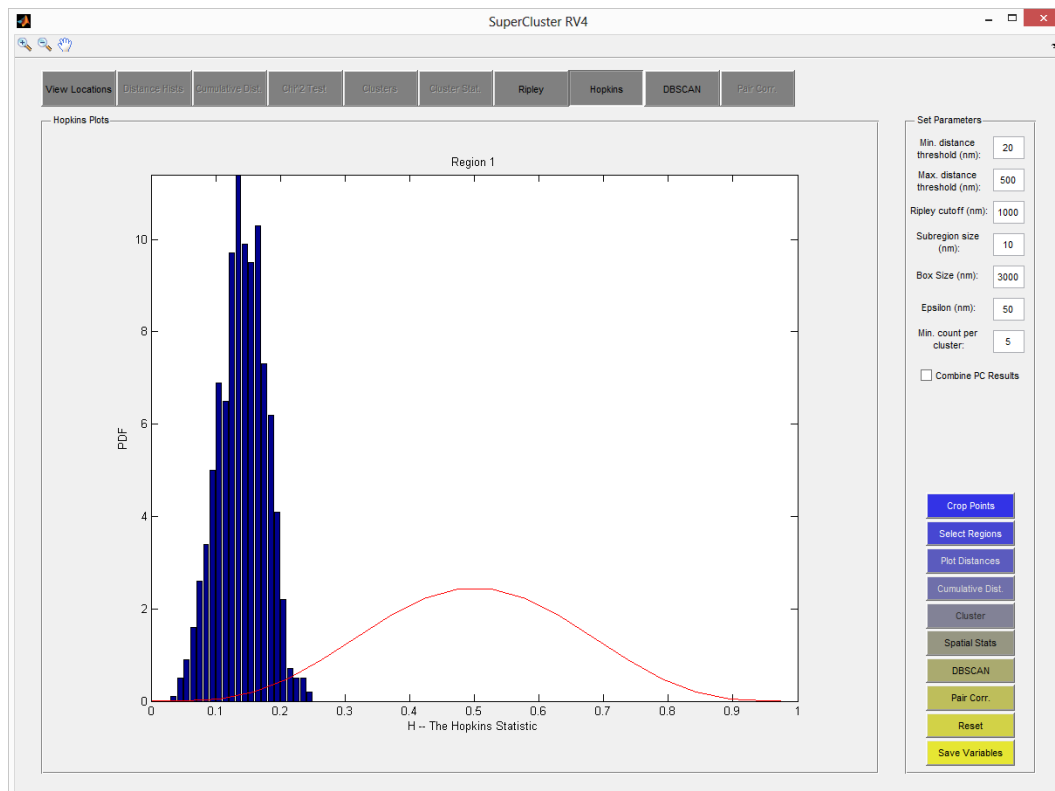


Figura 5 – Estatística de Hopkins para amostra uniformemente distribuída no espaço ($H \approx 0$)

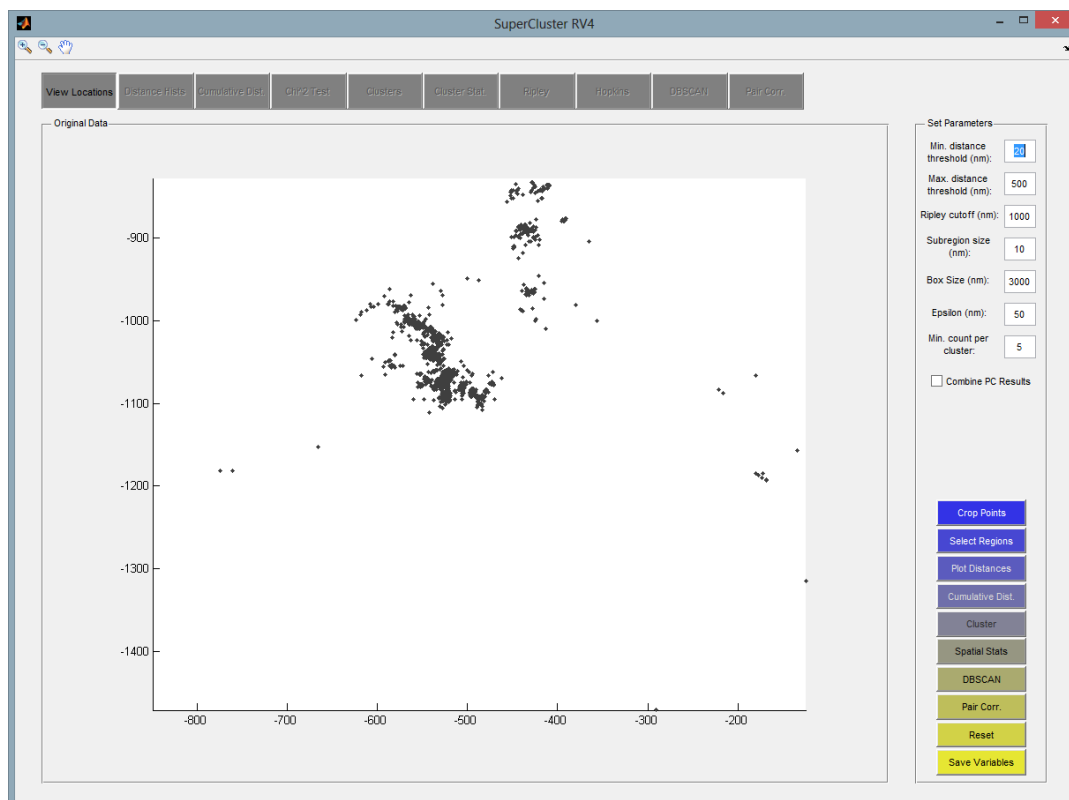


Figura 6 – Amostra com dados agrupáveis

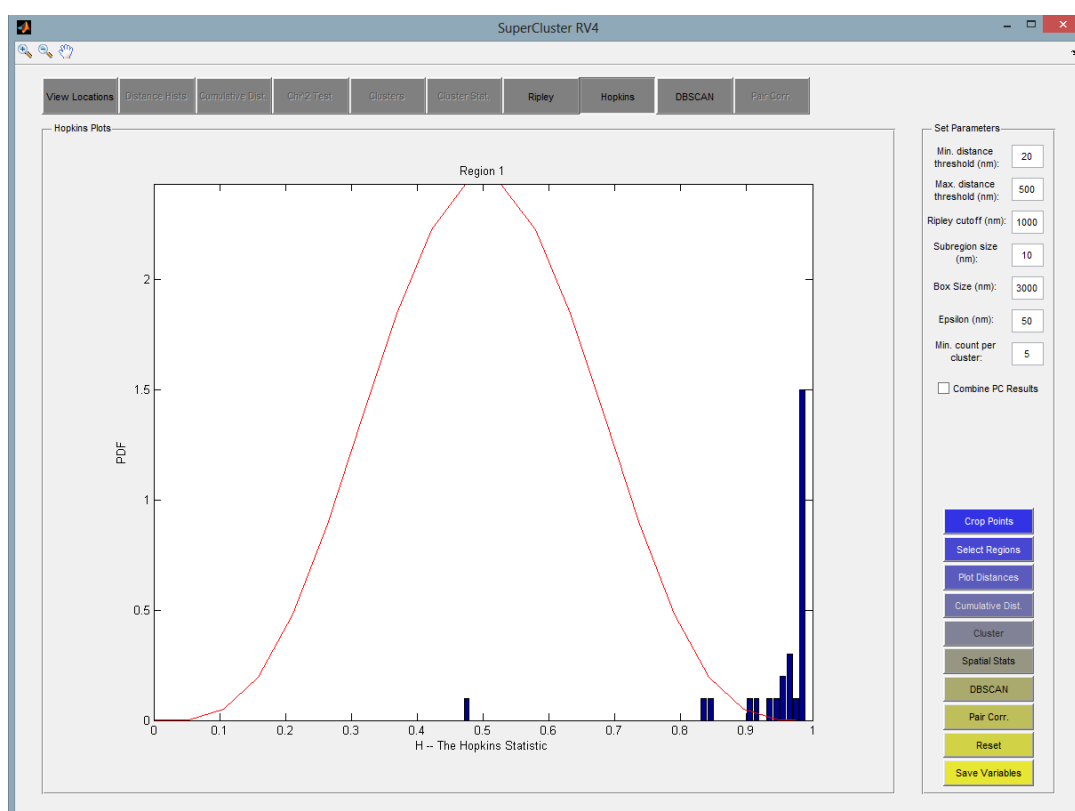


Figura 7 – Estatística de Hopkins para amostra com dados agrupáveis ($H \approx 1$)

ANEXO III.**Agrupamento dos Dados de Descargas Atmosféricas**

F. Metodologia de Agrupamento

Em mineração de dados as técnicas de agrupamento foram desenvolvidas com o intuito de identificar e agrupar objetos. Esta classificação é efetuada com base na análise da informação existente sobre esses objetos. Usando este tipo de técnicas é possível criar conjuntos de objetos que de alguma forma contenham similaridades entre si. Ao longo do tempo foram surgindo diferentes algoritmos de agrupamento com o objetivo de aumentar a qualidade dos resultados obtidos. Existem diferentes tipos de modelos de agrupamentos, como por exemplo: modelos de conectividade, modelos centrados, modelos distribuídos, modelos de densidade, entre outros [13].

Os principais métodos de agrupamento existentes na literatura podem ser categorizados da seguinte forma:

- Métodos baseados em particionamento

Esses métodos consistem na maneira mais fundamental de agrupamento. Eles recebem como entrada, além de um conjunto de dados D , um inteiro k que representa o número de grupos que se deseja encontrar. Inicialmente, o centro de cada grupo, que não necessariamente é um ponto de D , é determinado. A cada passo, os elementos de D são associados ao cluster definido pelo centro mais próximo de cada ponto. A distância entre um elemento de D e um centro é calculada de acordo com algum critério de similaridade estabelecido pelo algoritmo. Após a etapa de associação, os centros dos grupos são atualizados. Esse processo é repetido até que um ótimo local seja atingido. Dentre os algoritmos mais conhecidos nessa categoria destacam-se o k -means e o k -medoids.

No algoritmo k -means os centros dos grupos são conhecidos como centróides. A escolha dos centróides é comumente feita através de amostras aleatórias de D . Uma vez finalizada a etapa de associação de cada elemento do conjunto D ao seu centróide mais próximo, os valores dos centróides são

atualizados de acordo com a média dos elementos que estão associados àquele centróide. Esses passos são repetidos até que não haja modificação dos centróides.

A dificuldade da escolha dos valores iniciais para os centróides, além do número de centroides, são as principais desvantagens do algoritmo k-means. Adicionalmente, não existe a garantia da convergência do algoritmo para um ótimo global, tornando o resultado do agrupamento bastante sensível à inicialização dos centróides. Por outro lado, a baixa complexidade computacional do algoritmo descrito o torna atrativo para aplicações onde existe um conhecimento prévio sobre o domínio do conjunto de dados.

- Métodos hierárquicos

Os métodos hierárquicos têm como resultado um aninhamento de grupos e o grau de similaridade entre esses grupos. Os algoritmos hierárquicos podem ser subcategorizados em aglomerativos e divisivos. Os aglomerativos funcionam assumindo inicialmente que cada elemento do conjunto de dados representa um grupo. A partir daí os grupos se unem, de acordo com algum critério de similaridade. Já os divisivos consideram todo o conjunto de dados como um só grupo, que é dividido de maneira recursiva de acordo com a medida de similaridade estabelecida.

- Métodos baseados em densidade

Os métodos baseados em densidade assumem que os elementos que pertencem a um determinado grupo seguem uma mesma distribuição. O objetivo dessas técnicas é identificar os grupos de acordo com os parâmetros que descrevem essa distribuição. Em outras palavras, os grupos são modelados como regiões densas do conjunto de dados, divididos por áreas de regiões esparsas. Dentre as vantagens desses métodos destaca-se a capacidade de identificar grupos de formatos arbitrários, enquanto que os métodos baseados em particionamento apresentam melhores resultados com grupos de formato circular. Dentre os algoritmos baseados em densidade existentes na literatura, o DBSCAN se destaca como um dos mais utilizados.

- Métodos baseados em grade

Enquanto que os outros métodos de agrupamento discutidos são orientados aos dados, os métodos baseados em grade são orientados ao espaço. Essencialmente, o espaço é dividido em células de uma estrutura de grade independente da distribuição dos dados. Essa estrutura é capaz de representar os objetos do conjunto de entrada através das células da grade. Todas as operações de agrupamento são realizadas na estrutura de grade criada. Essa é uma das principais vantagens dessa classe de algoritmos, uma vez que seu desempenho não depende diretamente do número de entradas do conjunto de dados, mas sim do número de células em cada dimensão da grade.

Para escolher o algoritmo adequado, alguns fatores devem ser considerados:

- **Objetivo da aplicação:** sempre que a aplicação envolver a menor distância entre o centro do grupo em relação aos demais objetos, HAN [16] sugere que sejam usados os algoritmos k-means e k-medoids. Já em aplicações onde os dados espaciais são imagens de satélite ou mapas de vegetação, uso do solo, etc, é aconselhável o uso de algoritmos baseados em densidade, os quais criam grupos mais uniformes em termos de densidade, cor, pixels, e assim por diante.
- **Velocidade X qualidade:** o algoritmo de agrupamento ideal, para qualquer aplicação, deveria atender a dois critérios: velocidade de execução e qualidade dos clusters gerados [16]. O problema é que os algoritmos que geram grupos de boa qualidade são incapazes de ler grandes bases de dados. Para solucionar esse problema, HAN sugere que se comprima ou diminua o tamanho da base de dados e então se execute o algoritmo sobre a base condensada. O desafio, nesse caso, é reduzir a base sem que haja perda de informação importante a ponto de não prejudicar a geração dos grupos.
- **Características dos dados:** as características dos dados a serem agrupados devem ser consideradas durante a escolha do algoritmo.

- Domínio dos atributos: se os atributos dos objetos são numéricos, o agrupamento é facilitado. Se forem binários, categóricos ou ordinais, o agrupamento se complica, já que a maioria dos algoritmos que implementam essa técnica foi desenvolvida para o uso de atributos numéricos.
- Ruído: na escolha do algoritmo, deve ser observado o percentual de dados com ruído, pois alguns algoritmos são muito sensíveis ao ruído e isso pode afetar a qualidade do cluster gerado.

G.

Agrupamento Baseado em Densidade e DBSCAN

A maioria dos métodos Particionais e Hierárquicos agrupam objetos baseando-se na distância entre eles. Por isso, tais métodos podem encontrar dificuldades para descobrir grupos de formas distintas. Nos métodos de agrupamento baseados em densidade, grupos são definidos como regiões densas, separadas por regiões menos densas que representam ruídos. As regiões densas podem ter diferentes formas e os pontos dentro de uma região podem também estar distribuídos distintamente e, por isso, os métodos baseados em densidade são indicados para filtrar ruídos e descobrir grupos com forma distintas, tais como elíptica, cilíndrica, espiralada, etc., até os completamente cercados por outro grupo (HAN e KAMBER, 2001).

Nos métodos baseados em densidade, ESTER et al. (1996) atentam que ao observar um conjunto de objetos tais como os da Figura 15 a seguir, pode-se detectar grupos circulares no conjunto 1, grupos de formatos distintos no conjunto 2 e grupos de objetos e ruídos não pertencentes a qualquer dos grupos no conjunto 3. Este reconhecimento automático existe porque sem saber, o cérebro humano reconhece visualmente, neste caso bidimensional, que dentro de cada grupo existe uma densidade de objetos típica que é consideravelmente maior do que fora dos agrupamentos. Além disso, a densidade de áreas de ruído é menor do que a densidade em qualquer dos grupos. Involuntariamente, o cérebro humano entende a formação dos grupos usando o conceito de grupos formados por densidade para reconhecer os agrupamentos e ruídos nos exemplos mostrados na Figura 8. Um método baseado em densidade agrupa objetos baseado nesta noção de densidade e capta este comportamento, observados de forma visualmente óbvia nos exemplos acima, em conjuntos de dados de maiores dimensões onde a compreensão visual humana não é capaz de trabalhar.



Figura 8 – Conjunto de objetos com agrupamentos visualmente não globulares (conjunto 2) e presença de ruídos (conjunto 3).

No ano de 1996, Martin Ester, Hans-Peter Kriegel, Jörg Sander e Xiaowei Xu publicaram o artigo intitulado *A Density Based Spatial Clustering of Applications With Noise*. Nesse artigo foi apresentado o algoritmo DBSCAN, que pode ser aplicado a grandes conjuntos de dados que possuem ruídos (*outliers*), ao mesmo tempo em que encontra grupos com diversos formatos com eficiência aceitável.

O DBSCAN encontra agrupamentos baseando-se na vizinhança dos objetos, onde a densidade associada a um ponto é obtida por meio da contagem do número de pontos vizinhos em uma determinada região ao redor desse ponto [ERTÖZ, STEINBACH; KUMAR, 2003]. Esse algoritmo possui a capacidade de encontrar agrupamentos considerando as propriedades dos dados, pois não requer que seja informado antecipadamente o número de grupos, permitindo a formação de agrupamento com de diferentes formas. Em contrapartida, são necessários outros dois parâmetros de entrada para o algoritmo. Outras características importantes do algoritmo são a capacidade de identificar ruídos (*outliers*) e a possibilidade de poder trabalhar com diversas funções de distância [ANKERST et al, 1999; METZ, 2006; ESTER et al, 1996].

Os dois parâmetros de entrada que o DBSCAN necessita são:

a) raio de ϵ -vizinhança de um ponto: determina o raio de vizinhança ϵ para cada ponto da base de dados. Dado o parâmetro ϵ , o algoritmo DBSCAN verifica a quantidade de pontos contidos no raio ϵ para cada ponto da base e dados, e se essa quantidade exceder certo número, um cluster é formado;

b) número mínimo de pontos (η): parâmetro que especifica o número mínimo de pontos, no dado raio de ϵ -vizinhança, que um ponto precisa possuir para ser considerado um ponto central e consequentemente, de acordo com as definições de cluster baseado em densidade, iniciar a formação de um cluster.



Figura 9 – Agrupamentos de formatos arbitrários encontrados pelo algoritmo DBSCAN

Fonte: Data Mining - The Hypertextbook

A ideia principal do método DBSCAN é que para cada ponto de um agrupamento, a vizinhança para um dado raio contém, no mínimo, certo número de pontos, ou seja, a densidade na vizinhança tem que exceder um limiar. Para entender o método é necessário conhecer algumas definições específicas listadas a seguir.

Definição 1: (Vizinhança de um ponto) A vizinhança de um objeto p com raio ϵ é chamada de vizinhança de p é dada por: $N(p) = \{q \text{ em } D \mid \text{dist}(p, q) < \epsilon\}$. Na Figura 56 abaixo os círculos representam, respectivamente, a vizinhança do ponto q e vizinhança do ponto p .

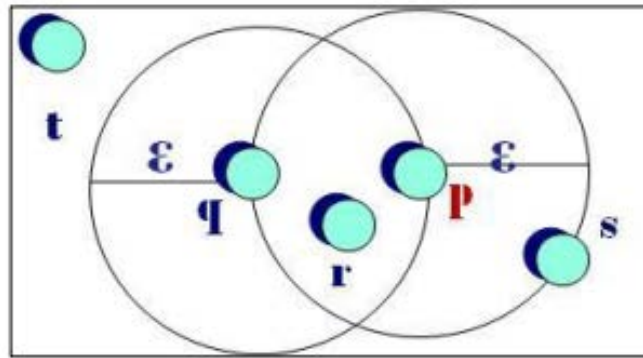


Figura 10 – ϵ -vizinhança de q e ϵ -vizinhança de p

Uma abordagem simples poderia exigir para cada ponto em um grupo que exista pelo menos um número mínimo (MinPts) de pontos na vizinhança daquele ponto. No entanto, esta abordagem falha porque há dois tipos de pontos em um cluster, pontos dentro do cluster (pontos centrais) e pontos na fronteira do cluster (pontos de borda).

Definição 2 (Ponto Central): Se a ϵ -vizinhança de um objeto p contém ao menos um número mínimo, MinPts, de objetos, então o objeto p é chamado de ponto central. Por exemplo, na Figura 5.1, se adotarmos MinPts = 4, p é um ponto central e os demais não são pontos centrais.

Definição 3 (pontos de borda): Se a ϵ -vizinhança de um objeto p contém menos que MinPts, mas contém algum ponto central, então o objeto p é chamado de ponto de borda. Na Figura 17, q, r e s são pontos de borda.

Em geral, a ϵ -vizinhança de um ponto de borda contém significativamente menos pontos do que a ϵ -vizinhança de um ponto central. Portanto, deve-se definir o número mínimo de pontos, MinPts para um valor relativamente baixo, de modo a incluir todos os pontos pertencentes a um mesmo cluster. Este valor, no entanto, não pode ser característica para o respectivo conjunto, particularmente na presença de ruído. Portanto, exige-se que, para cada ponto p em um cluster C exista um ponto q em C, de modo que p está dentro da ϵ -vizinhança de q e $N_\epsilon(q)$ contém pelo menos MinPts pontos. Esta definição é elaborada como segue:

Definição 4 (Alcance Direto por Densidade): Um objeto p é alcançável por densidade diretamente do objeto q , com respeito à ϵ e a MinPts, se p está na ϵ -vizinhança de q , e q é um ponto central.

Alcance direto por densidade é simétrico para pares de pontos centrais. Contudo, o alcance direto por densidade não é simétrico se um ponto central e um ponto de borda estão envolvidos. Na Figura 11 a seguir, por exemplo, p é alcançável por densidade diretamente de q , mas q não é alcançável por densidade diretamente de p porque q não é ponto central.

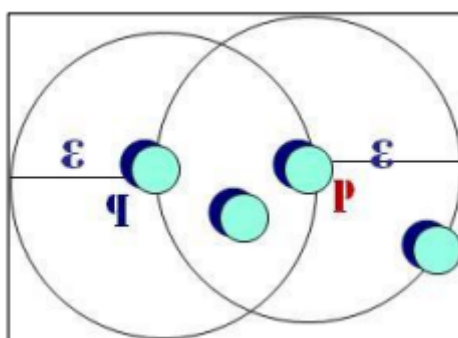


Figura 11 – ϵ -vizinhança de q e ϵ -vizinhança de p

Definição 5 (Alcance por Densidade): Um objeto p é alcançável por densidade do objeto q e MinPts em um conjunto D , se existe uma cadeia de objetos $\{p_1, \dots, p_n\}$, tais que $p_1 = q$ e $p_n = p$ e p_{i+1} é alcançável por densidade diretamente de p_i com respeito a ϵ e MinPts, para $1 \leq i \leq n$, p_i em D .

Na Figura 12, q é alcançável por densidade de p e p é alcançável por densidade de q . Portanto, existe um fechamento transitivo do alcance por densidade. Alcance por densidade é uma extensão canônica do alcance direto por densidade. Essa relação é transitiva, mas não é simétrica.

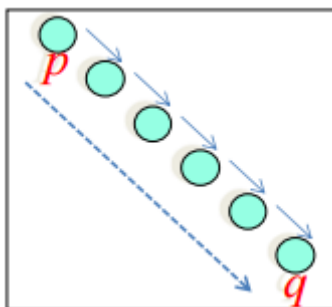


Figura 12 – Alcance por densidade no método DBSCAN

Embora não simétrica em geral, é óbvio que o alcance por densidade é simétrico para os pontos centrais. Dois pontos de fronteira de um mesmo cluster C não são possivelmente alcançáveis por densidade uns dos outros, porque a condição ponto central pode não valer para ambos. No entanto, deve haver um ponto central em C a partir do qual os dois pontos de fronteira de C são alcançáveis por densidade. Por isso, é introduzida a definição de densidade de conectividade por densidade que cobre esta relação de pontos de borda.

Definição 6 (Conexão por densidade): Um objeto p é conectado por densidade ao objeto q e MinPts em um conjunto de objetos D , se existe um objeto o em D tal que ambos p e q são alcançáveis por densidade do objeto o com respeito à MinPts. Na Figura 13 p e q são conectados por densidade através de o .

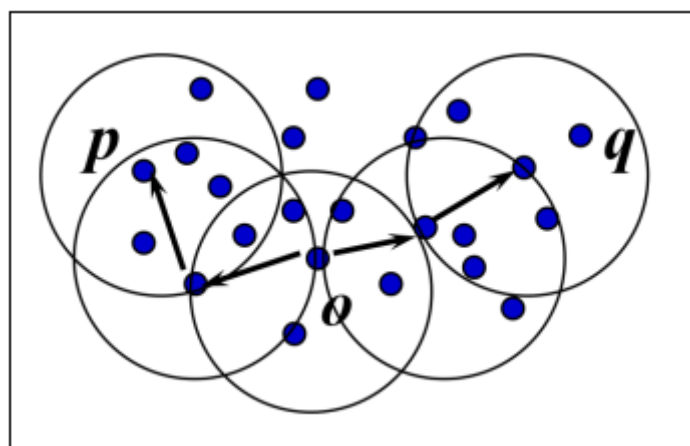


Figura 13 – Conexão por densidade no método DBSCAN

Conexão por densidade é uma relação simétrica. Para pontos alcançáveis por densidade a relação de conectividade por densidade é também reflexiva.

Definição 7 (Agrupamento DBSCAN): Seja D uma base de dados de pontos. Um grupo C com respeito à MinPts, sendo um subconjunto não vazio de D satisfazendo as seguintes condições:

- 1) $\forall p, q$: se $p \in C$ e q é alcançável por densidade a partir de p com respeito à MinPts, então $q \in C$ (Maximalidade).

2) $\forall p, q \in C$: p é conectado por densidade a q com respeito à MinPts (Conectividade).

Intuitivamente, um cluster DBSCAN é o conjunto de pontos conectados por densidade que é maximal com respeito ao alcance por densidade.

Definição 8 (Ruído): Sejam C_1, \dots, C_k os clusters da base de dados D com respeito aos parâmetros ϵ e MinPts, $i = 1, \dots, K$. Então, define-se o ruído como o conjunto de pontos na base de dados D que não pertença a qualquer grupo C_i , ou seja, o ruído = $\{p \in D \mid \forall i: p \notin C_i\}$. Um objeto que não é ponto central nem ponto de borda é ruído.

Assim definido, um cluster C com respeito à MinPts contém pelo menos MinPts pontos por causa das seguintes razões. Uma vez que C contém pelo menos um ponto p , p deve ser conectado por densidade a si mesmo, através de algum ponto o (que pode ser igual a p). Assim, pelo menos o deve satisfazer a condição de ponto central e, conseqüentemente, a ϵ -vizinhança de o contém pelo menos MinPts pontos.

De acordo com as definições 7 e 8, o DBSCAN é designado para descobrir clusters e ruído em uma base de dados espacial. Isso qualifica o método em classificar diretamente os ruídos, os quais seriam considerados de forma equivocada nos agrupamentos.

H.

Agrupamento dos Dados de Descargas Atmosféricas Baseado em Densidade

Conforme discutido anteriormente, dentre os parâmetros para estudos de desempenho de sistemas elétricos, a densidade de descargas para o solo é um dos mais importantes, pois determina a quantidade de descargas incidentes a partir do número de dias com ocorrência de trovoadas por ano, que é o nível ceráunico da região. Dessa forma, o desligamento de uma linha de transmissão está normalmente associado a agrupamentos de descargas atmosféricas, não sendo comum observar individualmente esses eventos como responsáveis por tais desligamentos.

Além dessas características, os agrupamentos de descargas atmosféricas possuem regiões densas e de formas distintas, com a presença de ruídos, os quais dificultam a identificação de agrupamentos através de métodos tradicionais, tais como K-Means e Redes Auto-Organizáveis de Kohonen [17]. Aliado a esses fatores, a dificuldade de conhecer previamente o número de agrupamentos presentes em eventos de descargas atmosféricas foi decisivo para a busca de algoritmos baseados em densidade, apropriados para agrupar dados com essas características identificadas.

O algoritmo DBSCAN (Density Based Spatial Clustering of Application with Noise (Clusterização Espacial Baseada em Densidade de Aplicações com Ruído) foi utilizado nesse estudo como o método de agrupamento dos dados, por ser capaz de identificar grupos de formatos distintos e separação eficiente de ruídos (*outliers*).

O algoritmo DBSCAN foi testado para descargas atmosféricas próximas a linhas de transmissão, apresentando bom desempenho no agrupamento dos eventos de descarga e na identificação de ruídos, sendo possível observar a sensibilidade no ajuste dos parâmetros de entrada Eps e MinPts. O estudo abaixo foi realizado sobre uma amostra de descargas atmosférica no dia 26/09/2014 às 04hs01min, próximos a linha de transmissão LT Cascavel Oeste/Ivaiporã C1, para diferentes valores de Eps e MinPts e observados em intervalos a cada 30 minutos.

No eixo x foi representada a longitude e no eixo y a latitude, sendo os agrupamentos diferenciados através de códigos de cores.

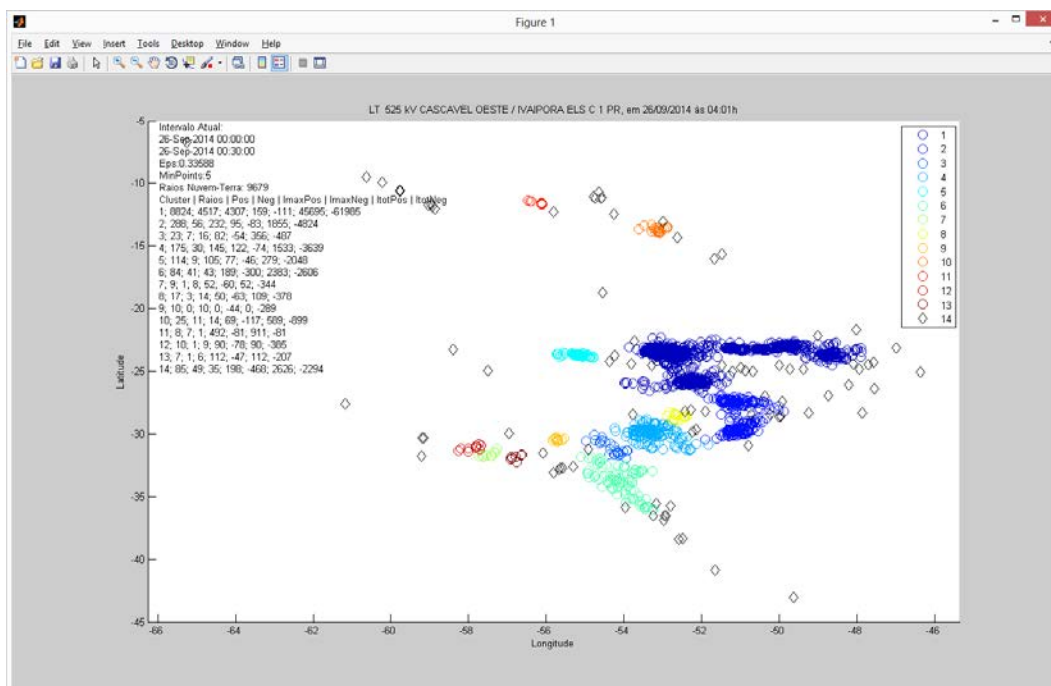


Figura 14 – DBSCAN para Eps=0,33588 e MinPts=5 (Intervalo 00:00:00 a 00:30:00)

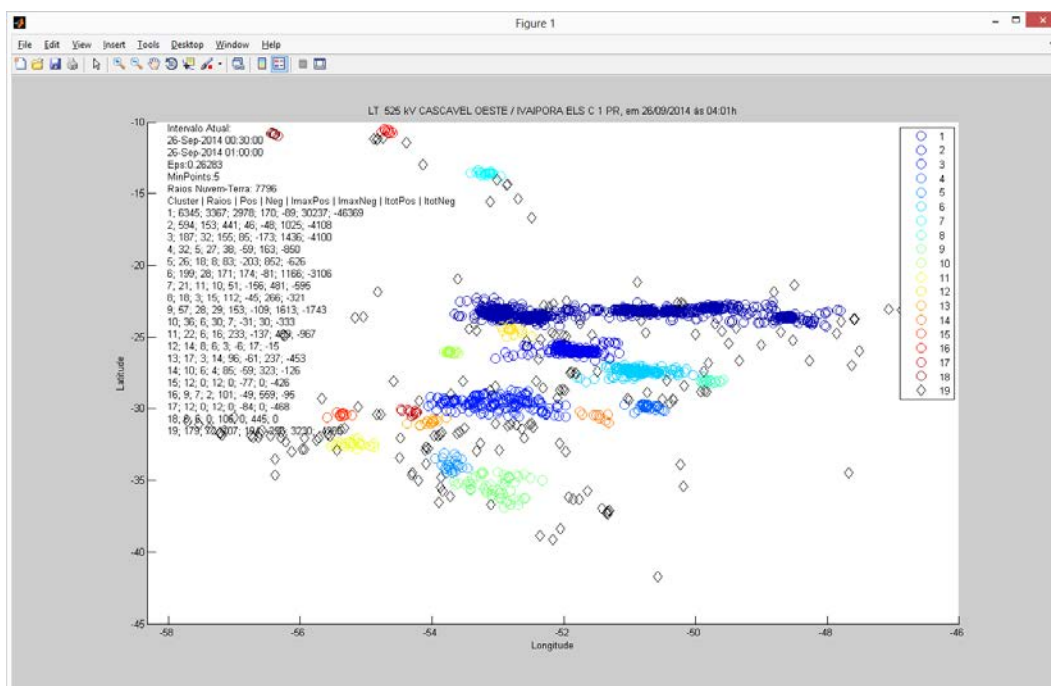


Figura 15– DBSCAN para Eps=0,26283 e MinPts=5 (Intervalo 00:30:00 a 01:00:00)

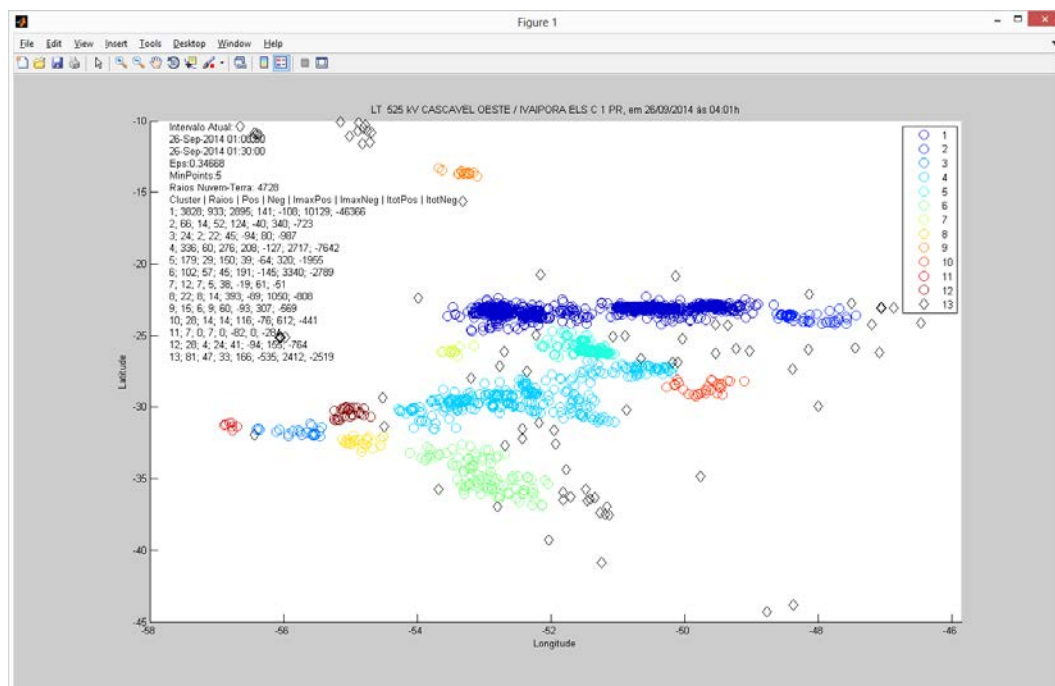


Figura 16 – DBSCAN para Eps=0,34688 e MinPts=5 (Intervalo 01:00:00 a 01:30:00)

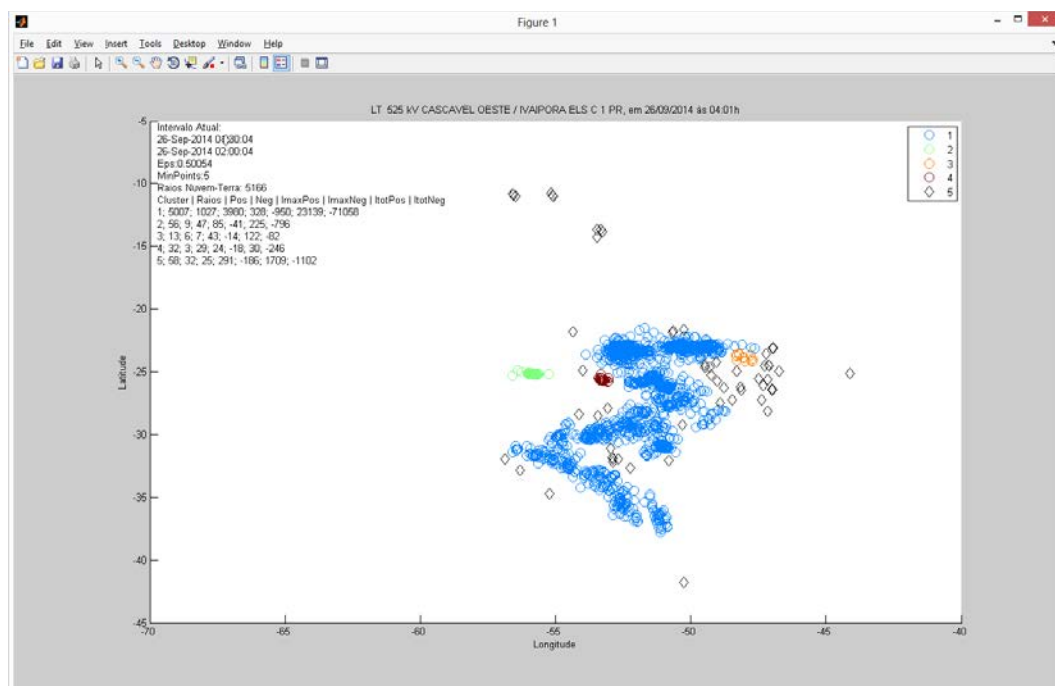


Figura 17 – DBSCAN para Eps=0,50054 e MinPts=5 (Intervalo 01:30:00 a 02:00:00)

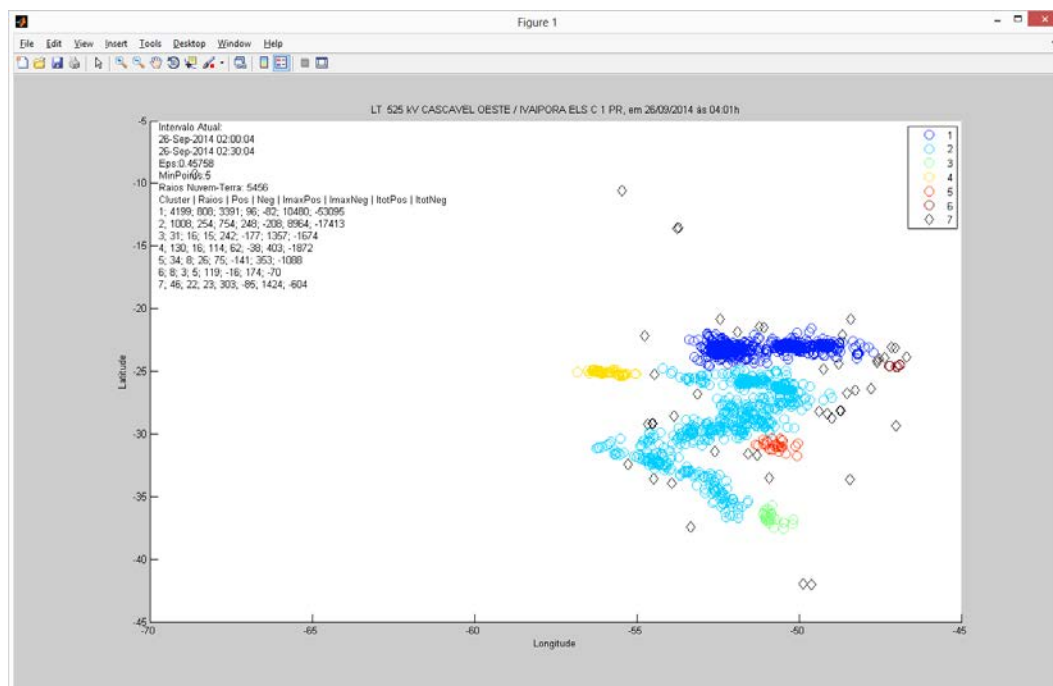


Figura 18 – DBSCAN para Eps=0,45758 e MinPts=5 (Intervalo 02:00:04 a 02:30:04)

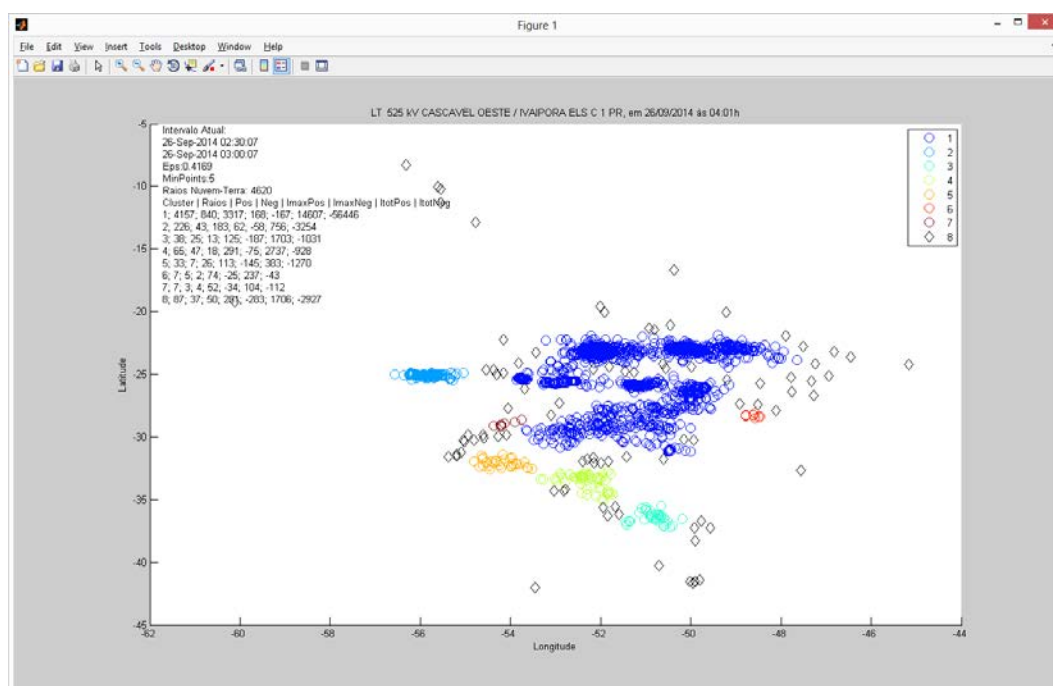


Figura 19 – DBSCAN para Eps=0,4169 e MinPts=5 (Intervalo 02:30:07 a 03:00:07)

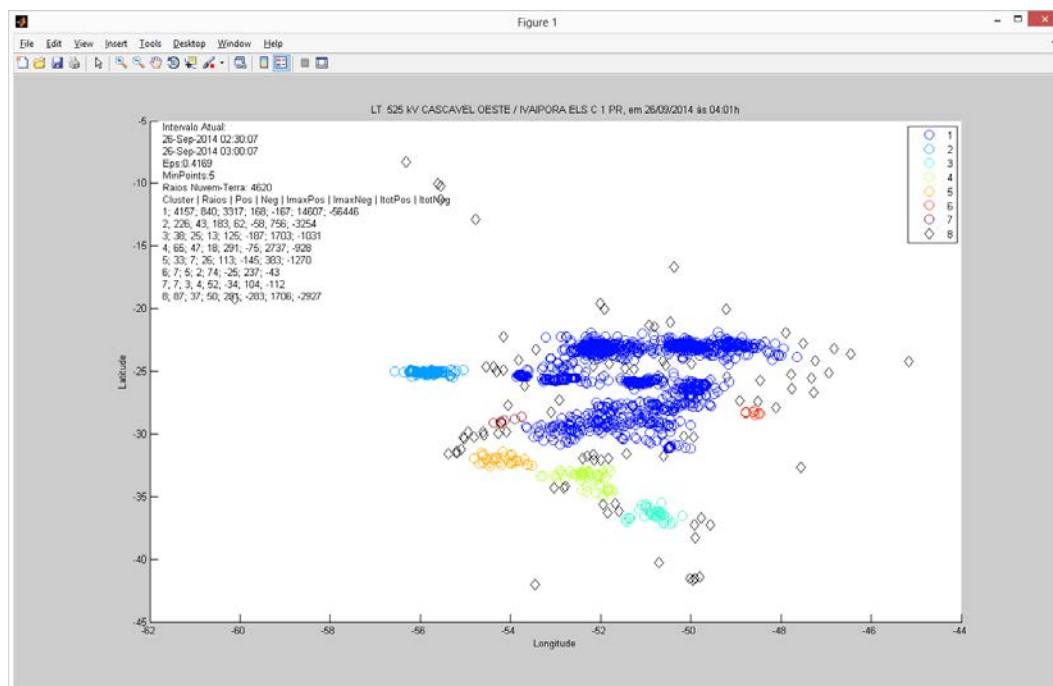


Figura 20 – DBSCAN para Eps=0,51598 e MinPts=5 (Intervalo 03:00:09 a 03:30:09)

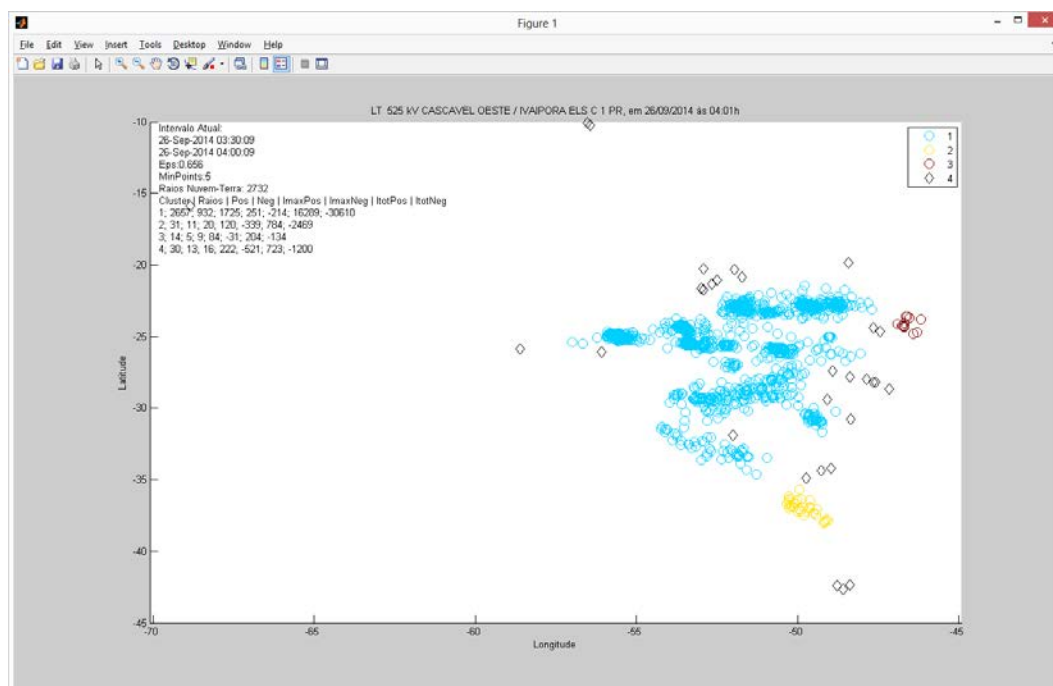


Figura 21 – DBSCAN para Eps=0,656 e MinPts=5 (Intervalo 03:30:09 a 04:00:09)

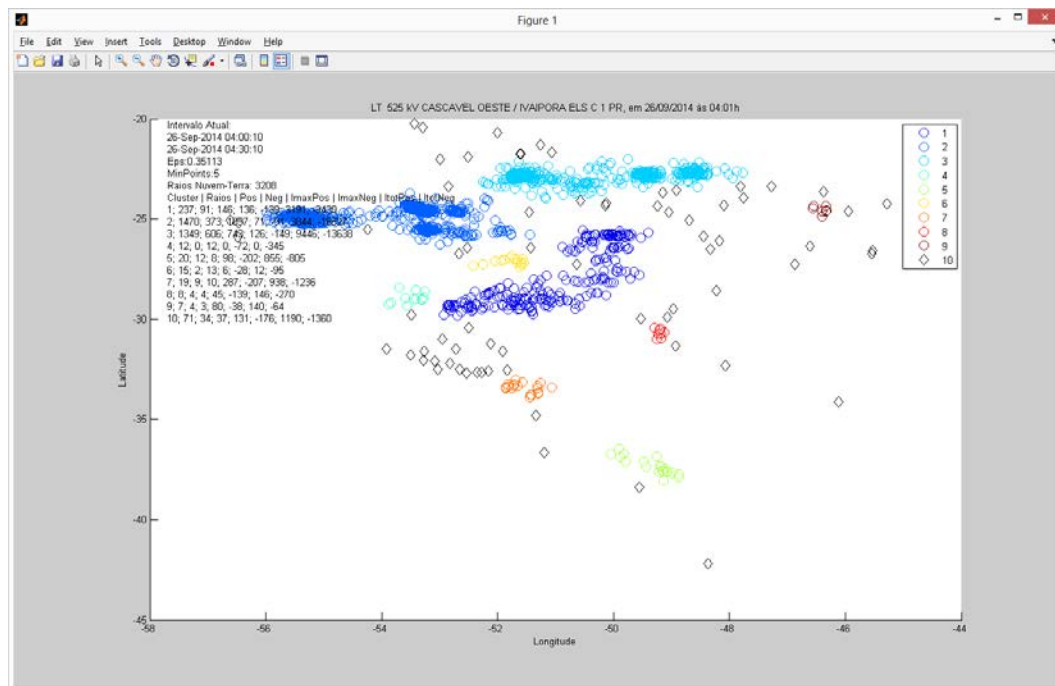


Figura 22 – DBSCAN para Eps=0,35113 e MinPts=5 (Intervalo 04:00:10 a 04:30:10)

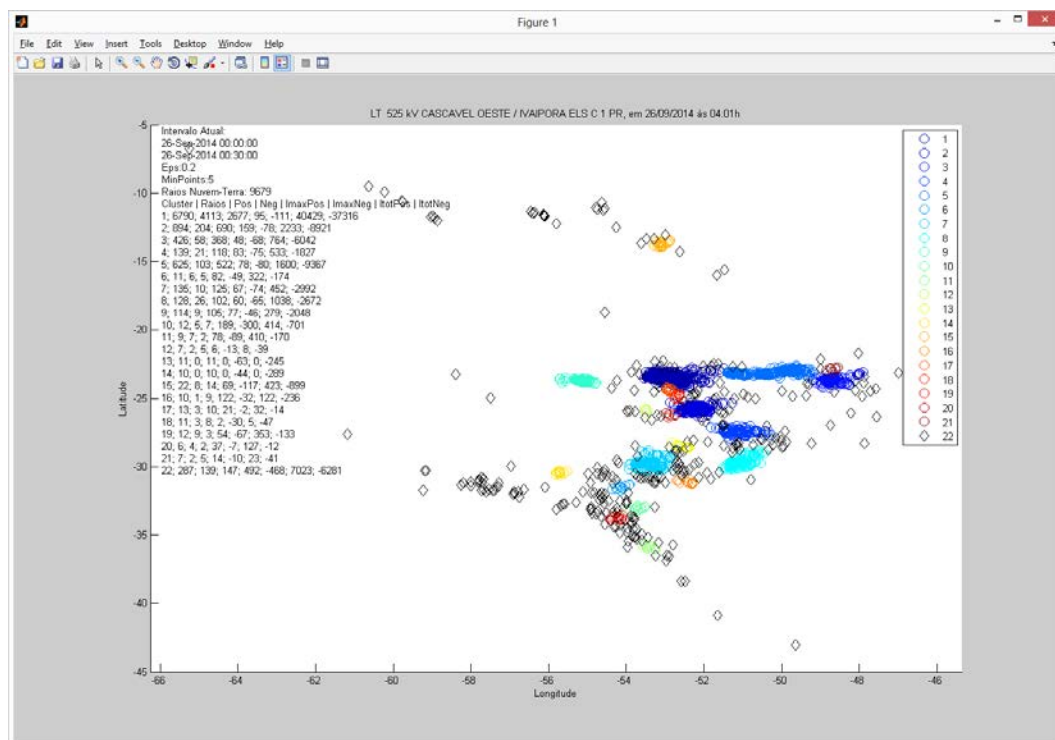


Figura 23 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 00:00:00 a 00:30:00)

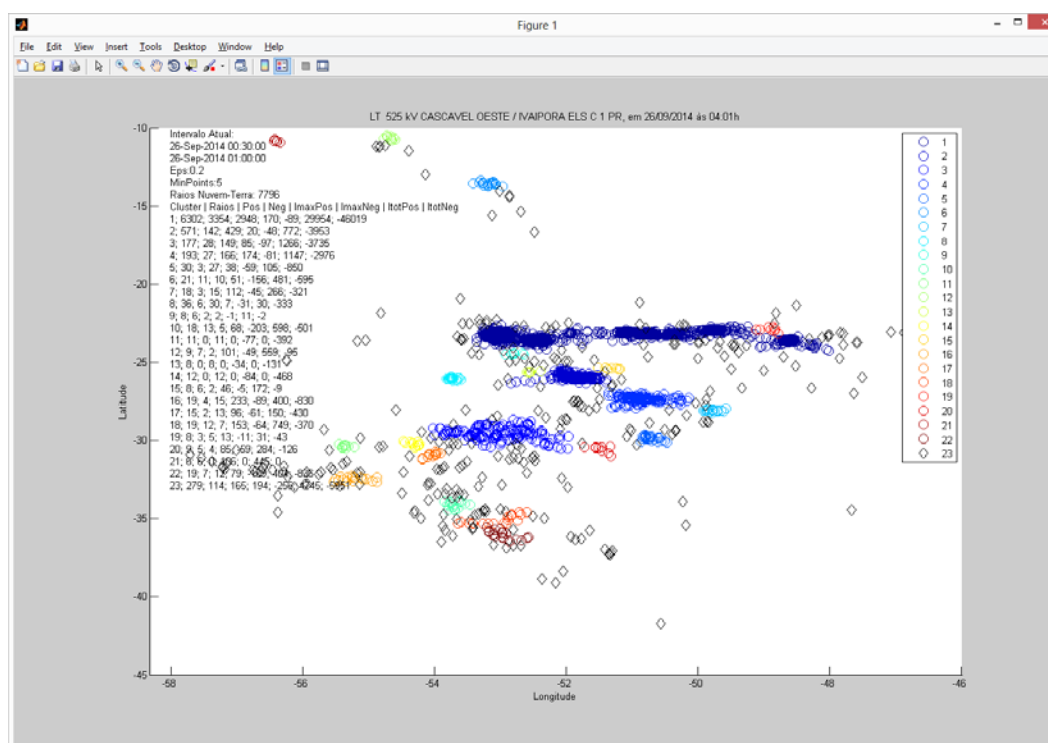


Figura 24 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 00:30:00 a 01:00:00)

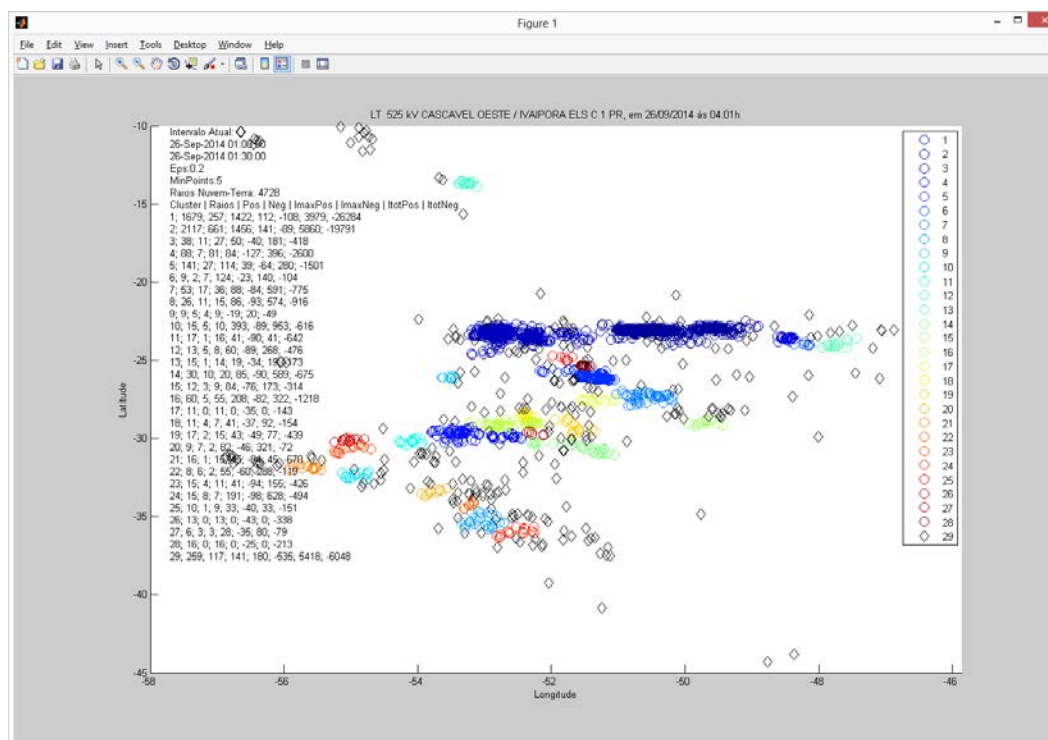


Figura 25 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 01:00:00 a 01:30:00)

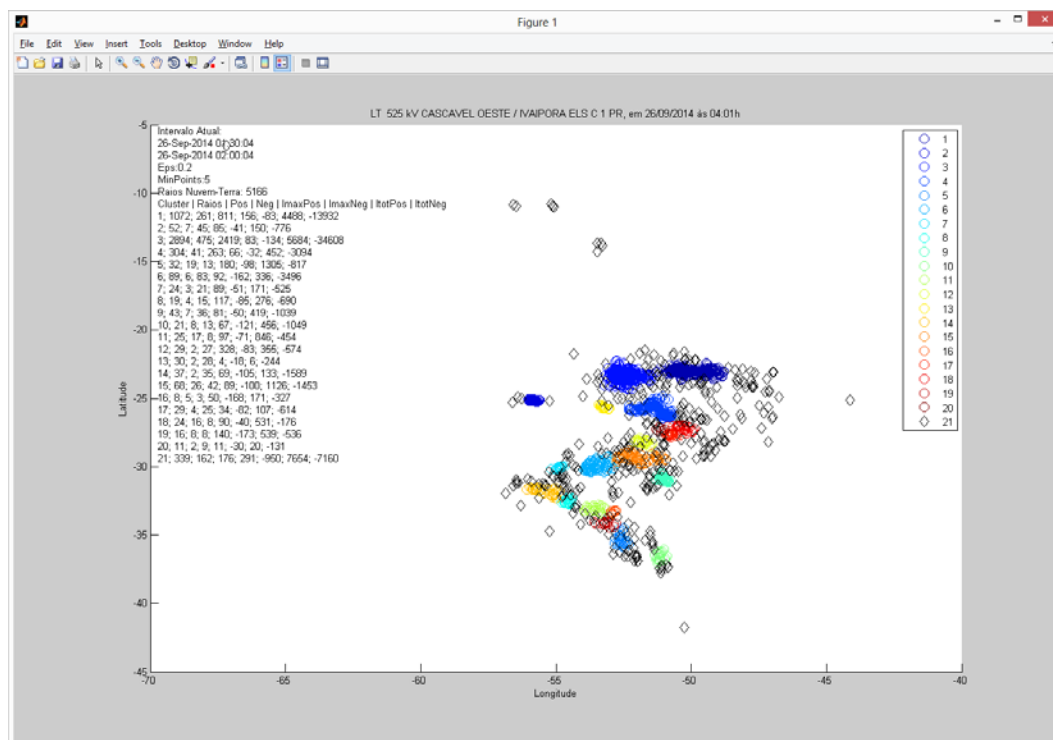


Figura 26 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 01:30:04 a 02:00:04)

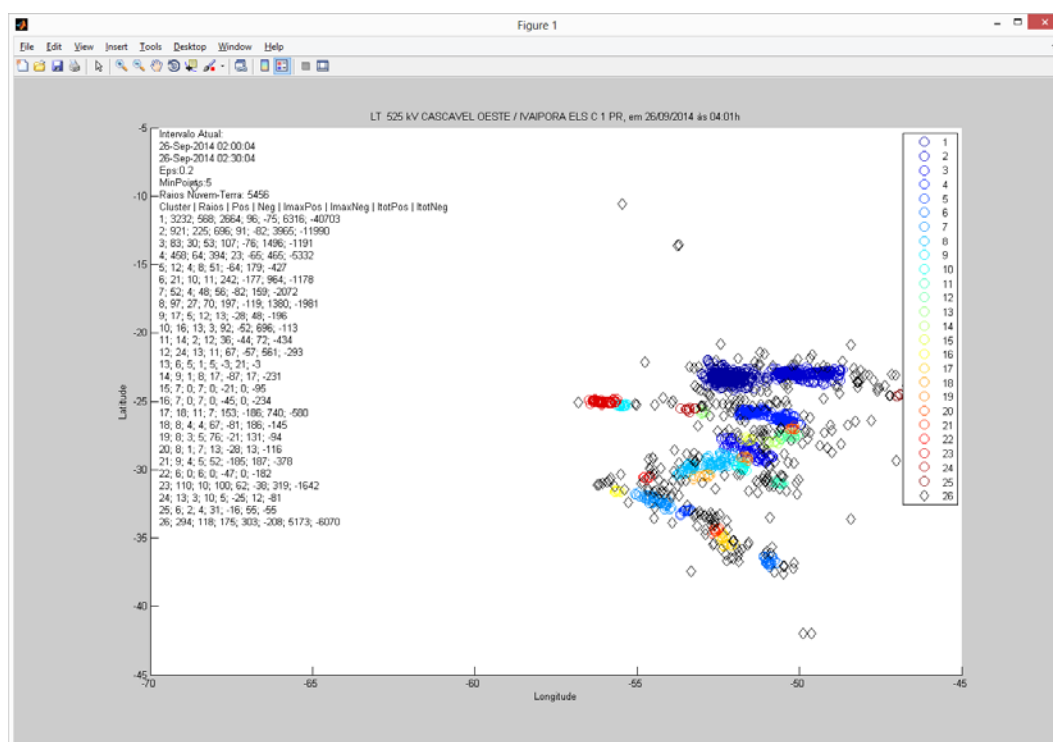


Figura 27 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 02:00:04 a 02:30:04)

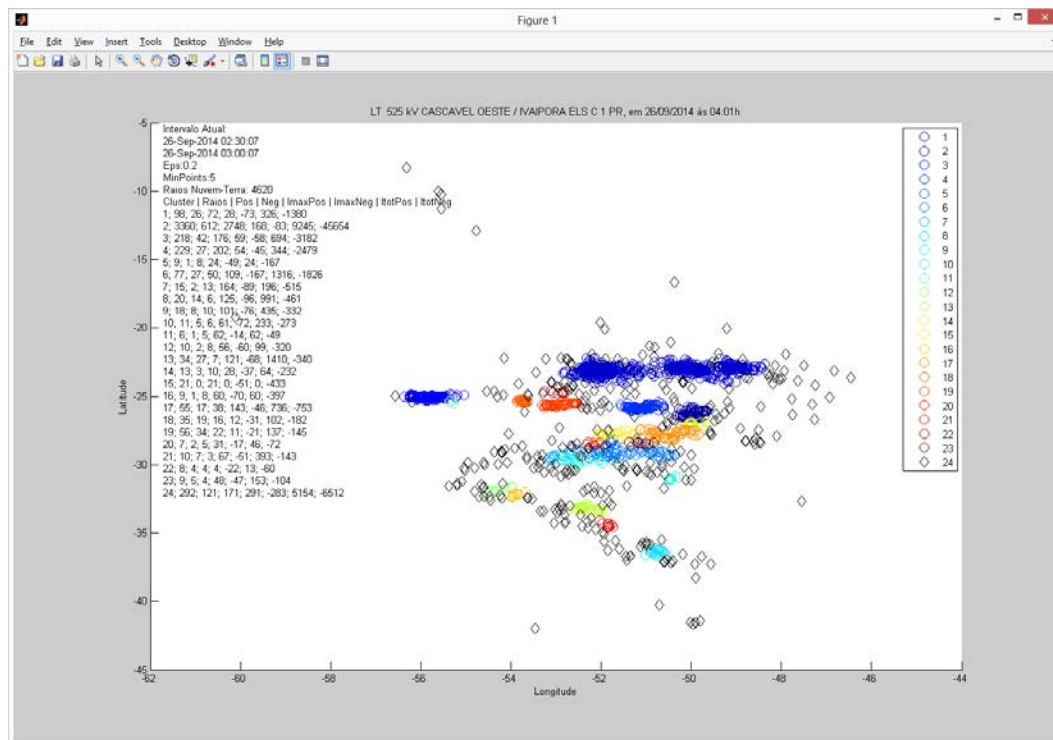


Figura 28 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 02:30:07 a 03:00:07)

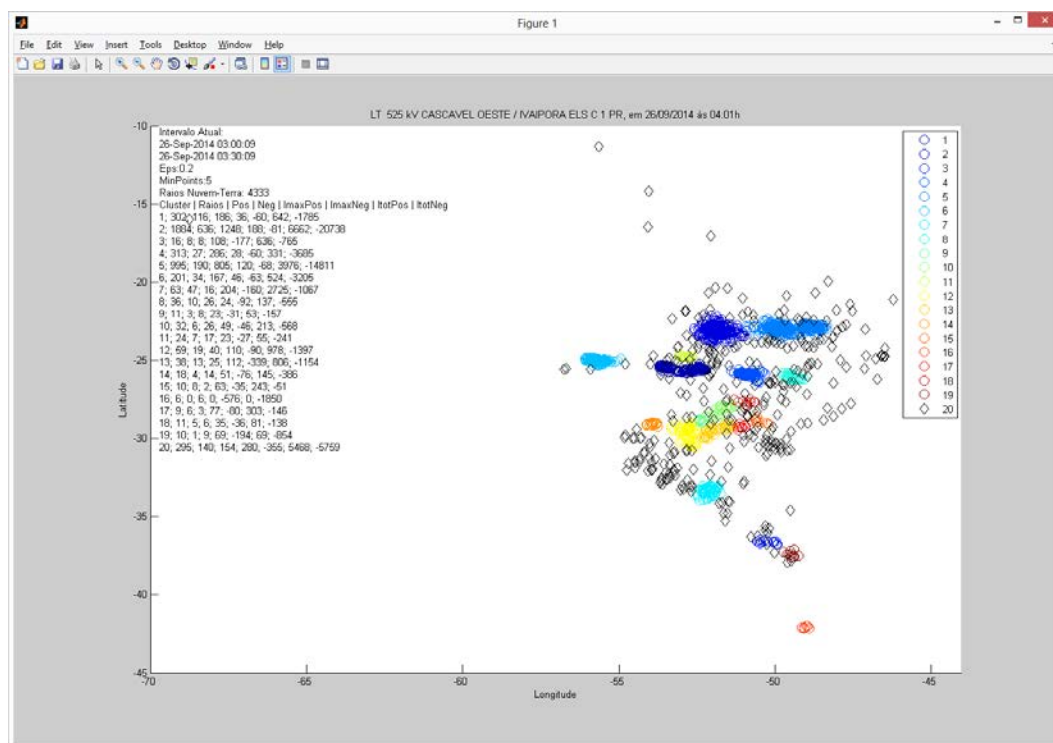


Figura 29 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 03:00:09 a 03:30:09)

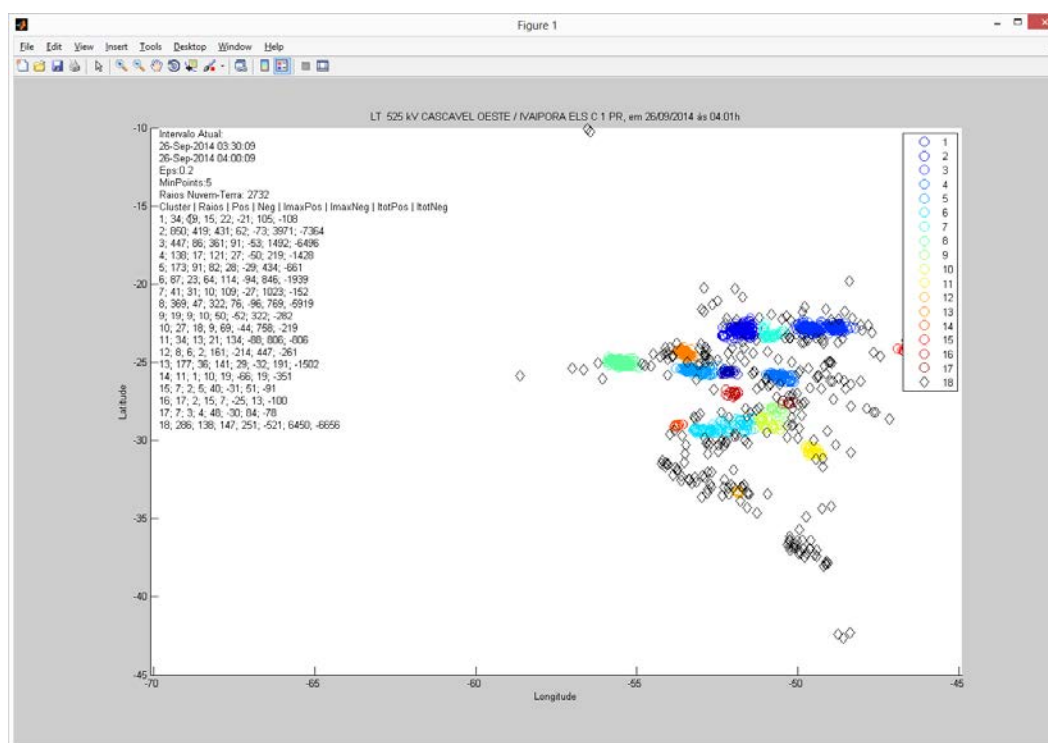


Figura 30 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 03:30:09 a 04:00:09)

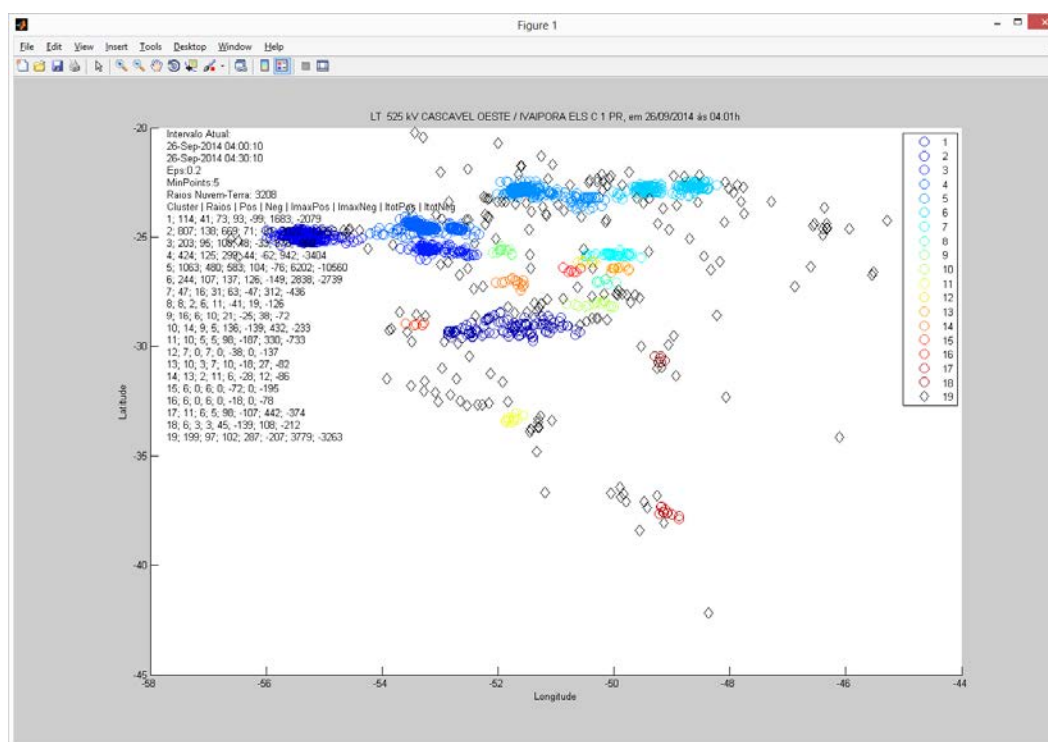


Figura 31 – DBSCAN para Eps=0,2 e MinPts=5 (Intervalo 04:00:10 a 04:30:10)

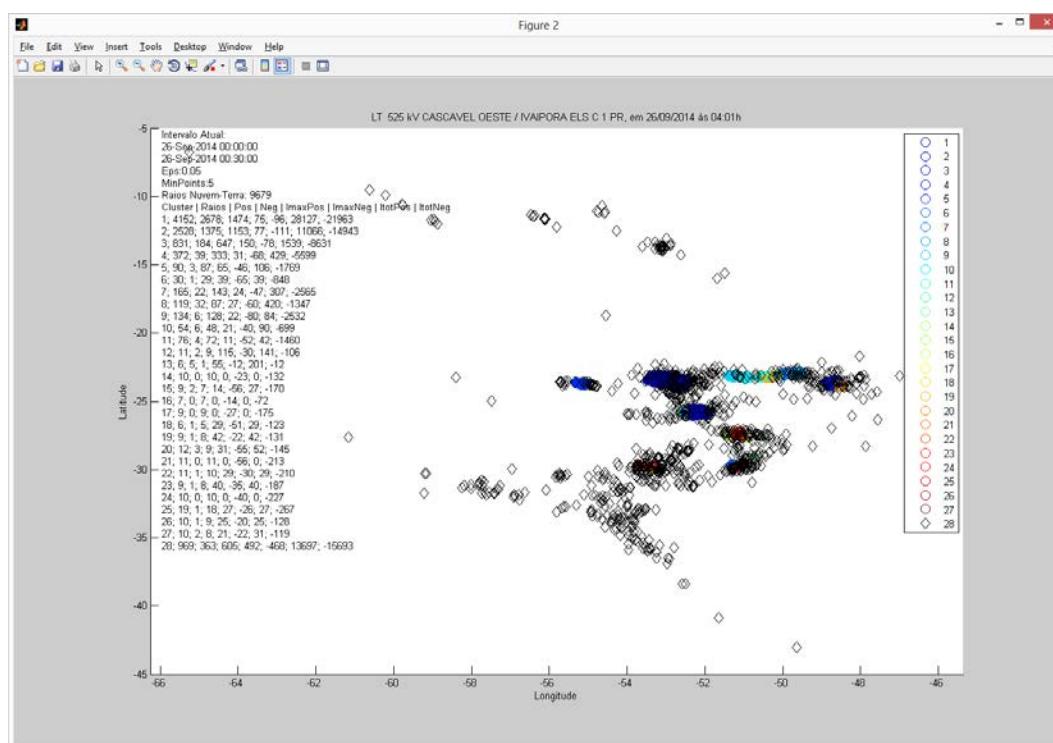


Figura 32 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 00:00:00 a 00:30:00)

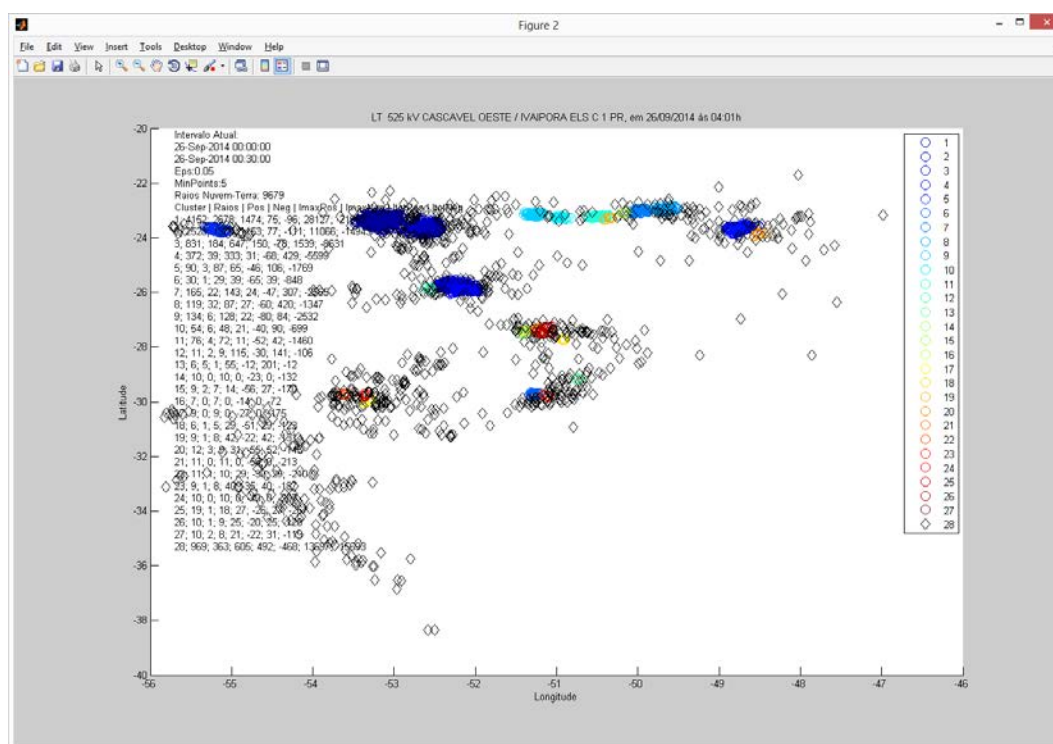


Figura 33 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 00:00:00 a 00:30:00)

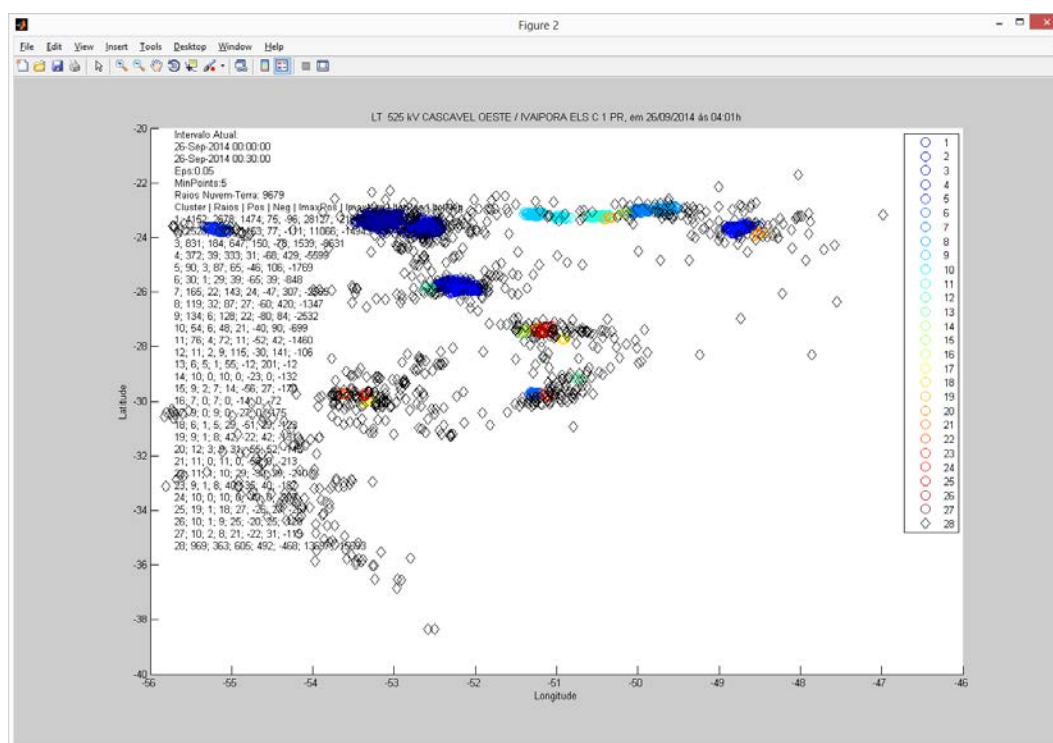


Figura 34 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 00:00:00 a 00:30:00)

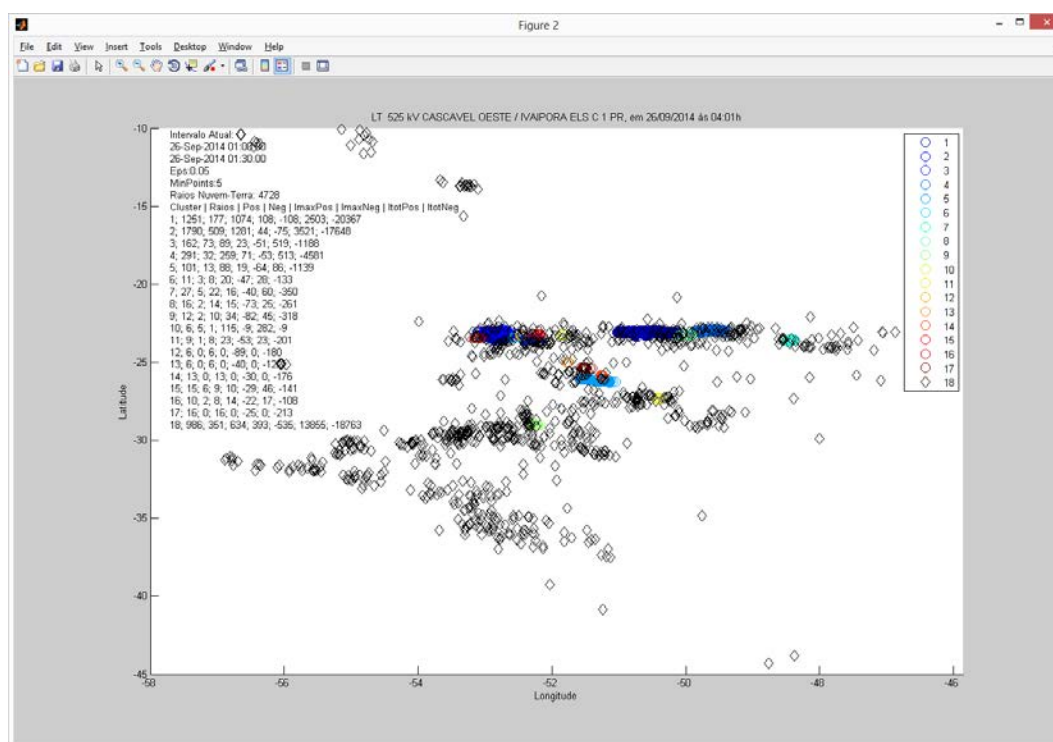


Figura 35 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 01:00:00 a 01:30:00)

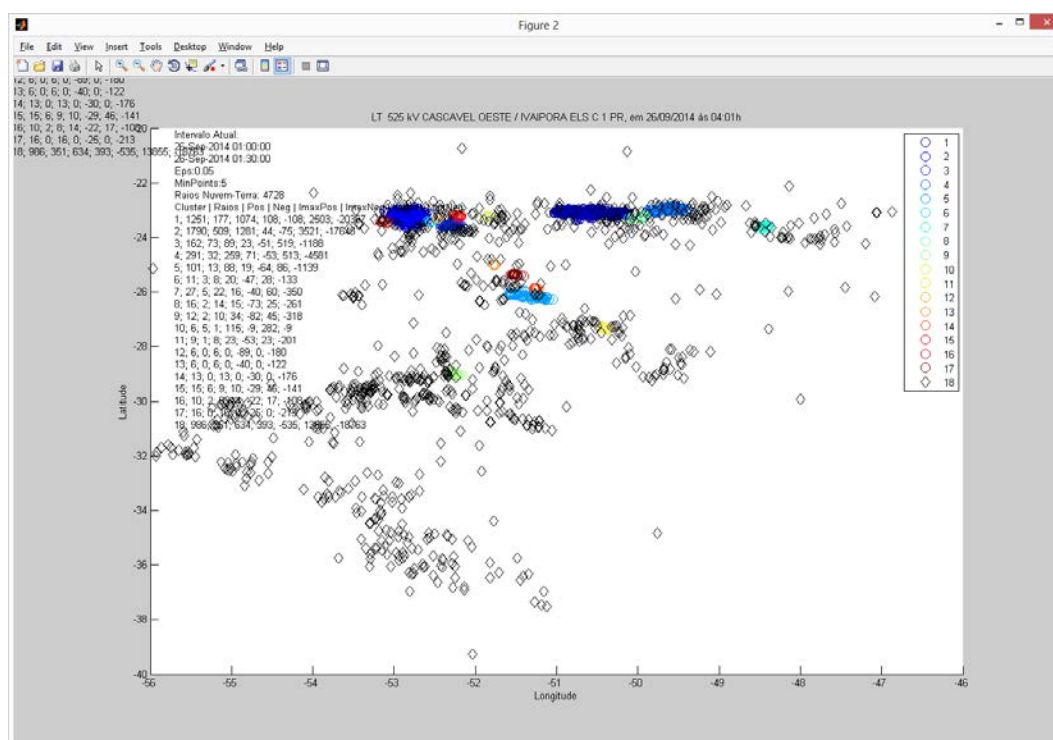


Figura 36 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 01:00:00 a 01:30:00)

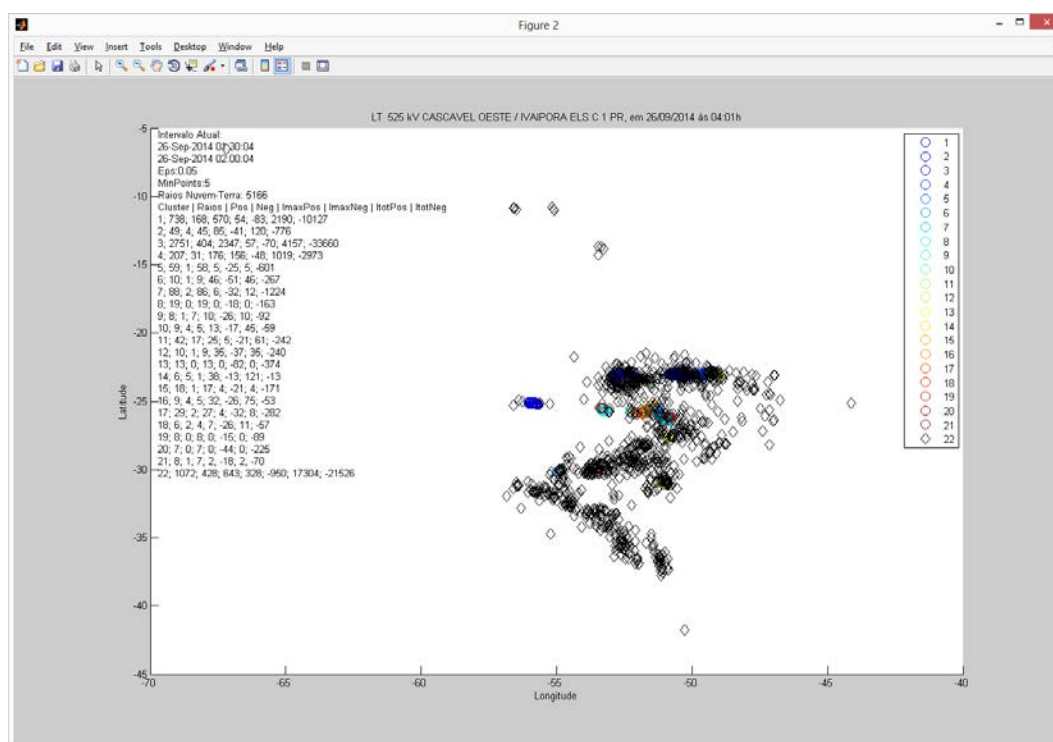


Figura 37 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 01:30:04 a 02:00:04)

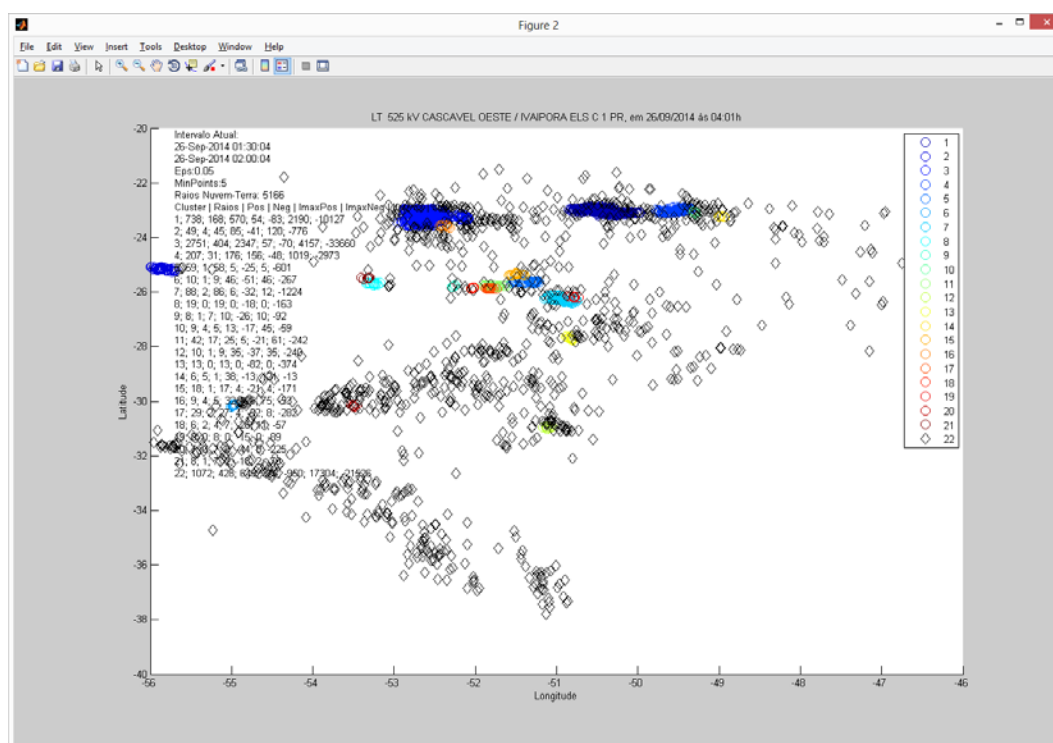


Figura 38 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 01:30:04 a 02:00:04)

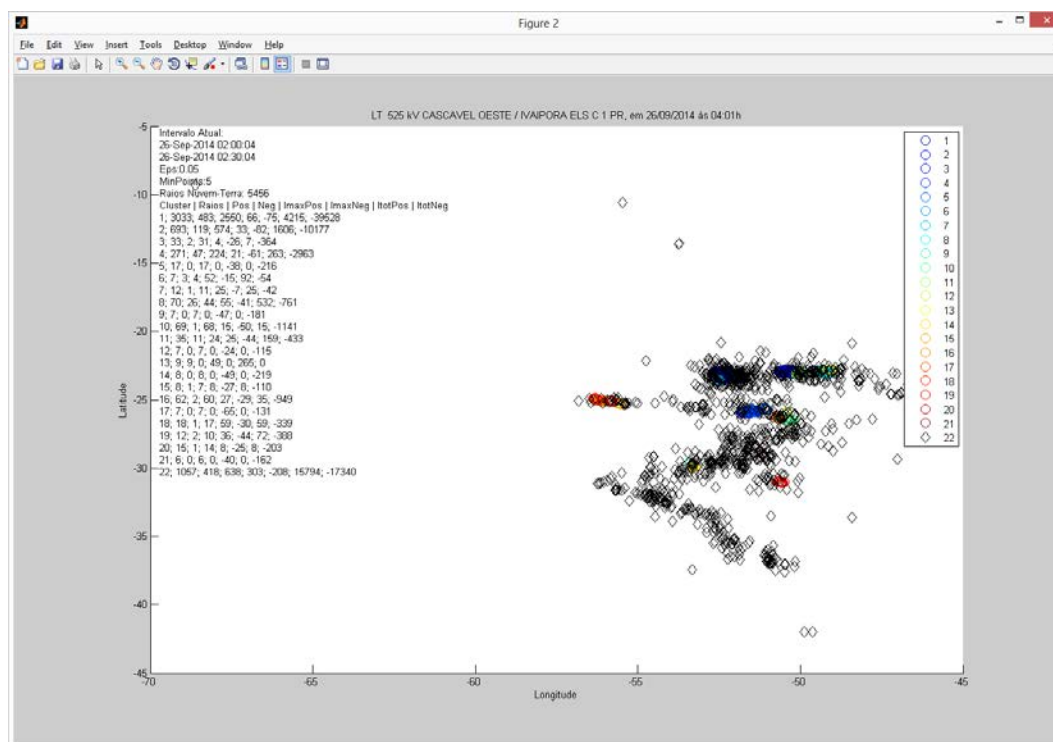


Figura 39 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 02:00:04 a 02:30:04)

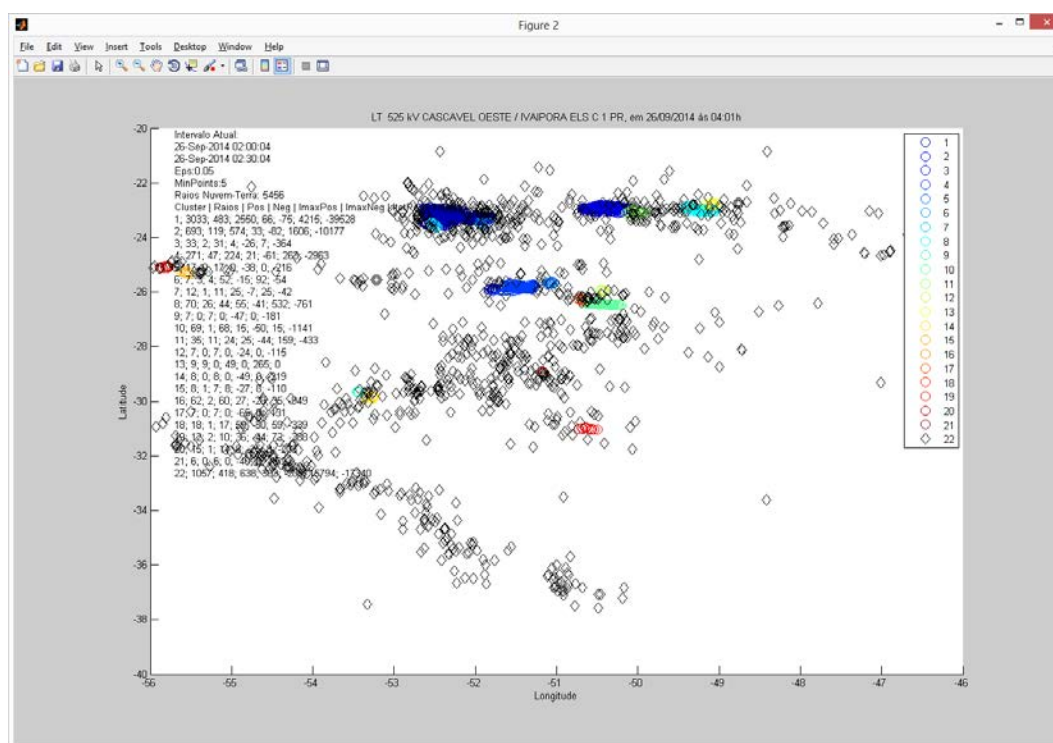


Figura 40 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 02:00:04 a 02:30:04)

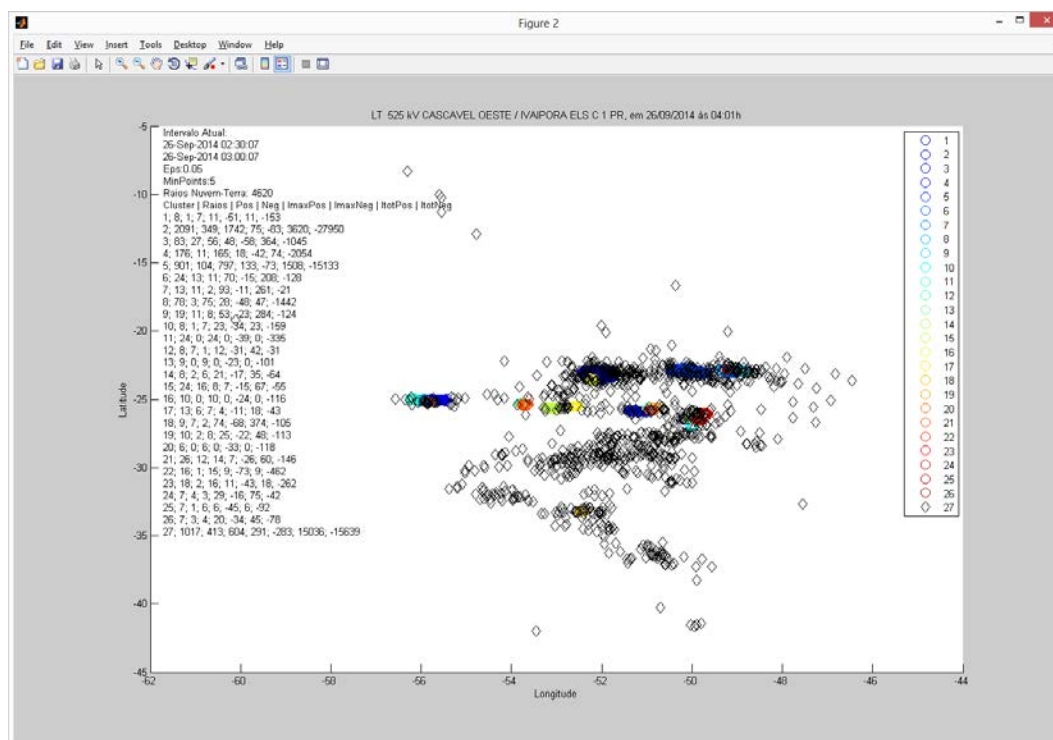


Figura 41 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 02:30:07 a 03:00:07)

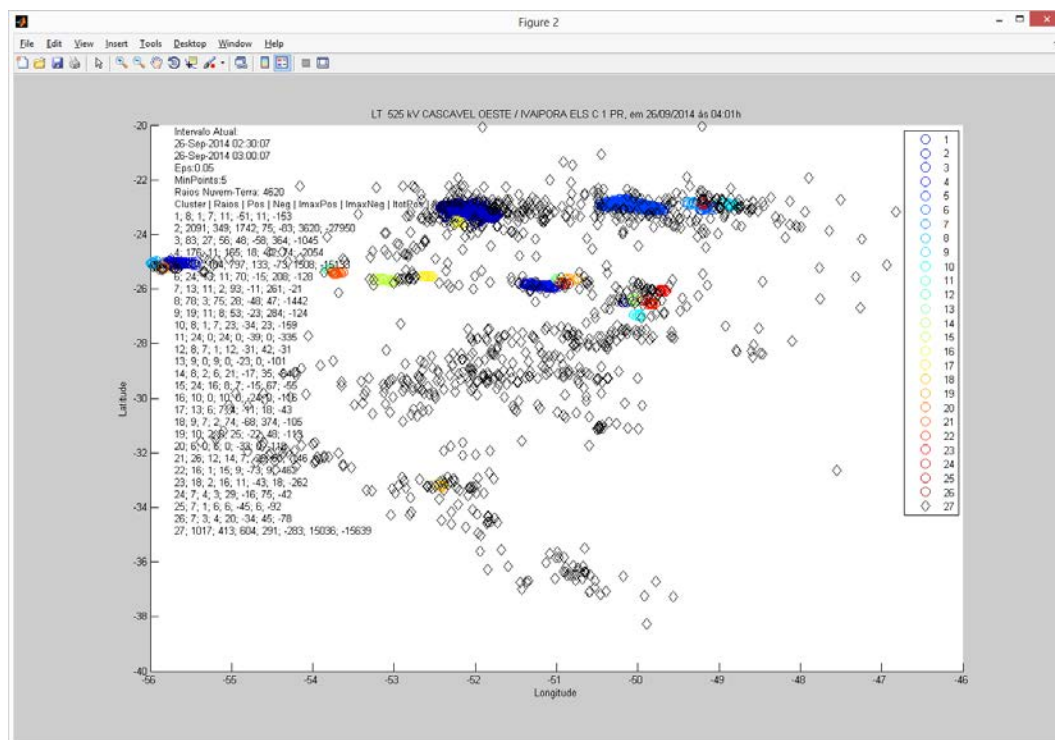


Figura 42 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 02:30:07 a 03:00:07)

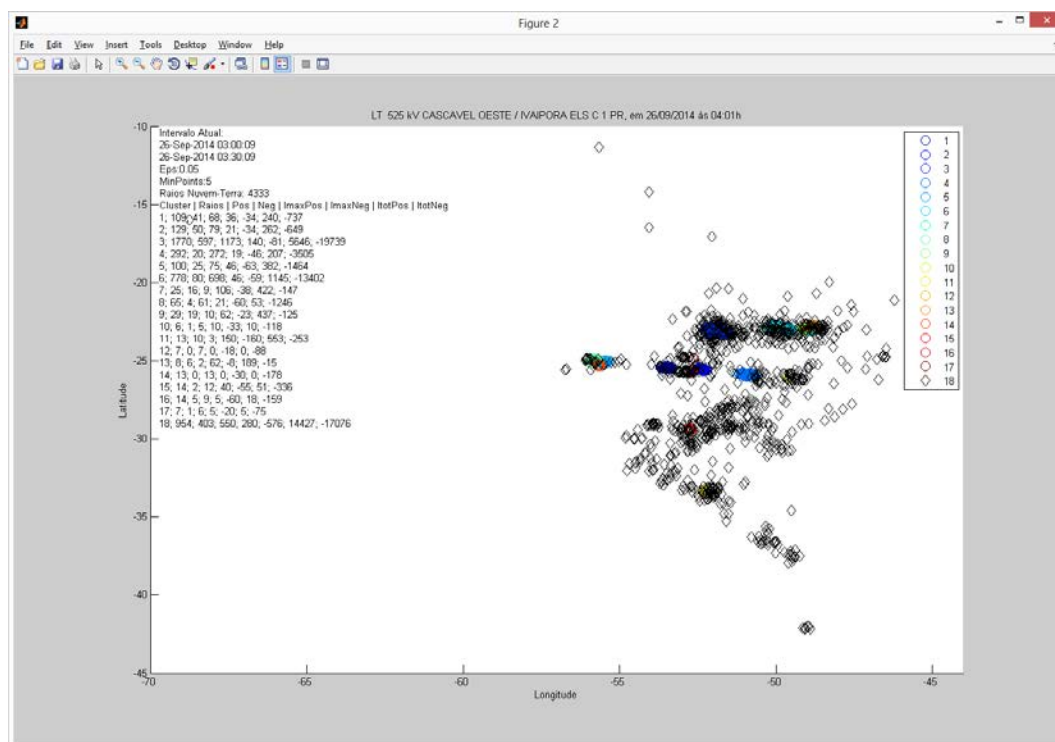


Figura 43 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 03:00:09 a 03:30:09)

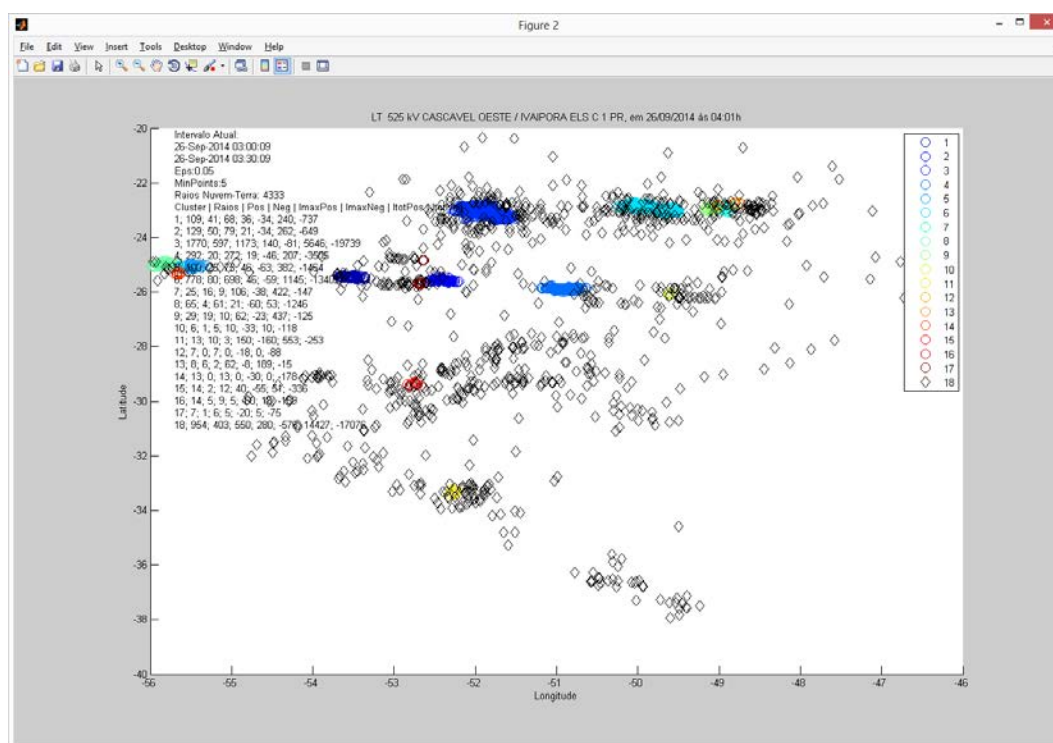


Figura 44 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 03:00:09 a 03:30:09)

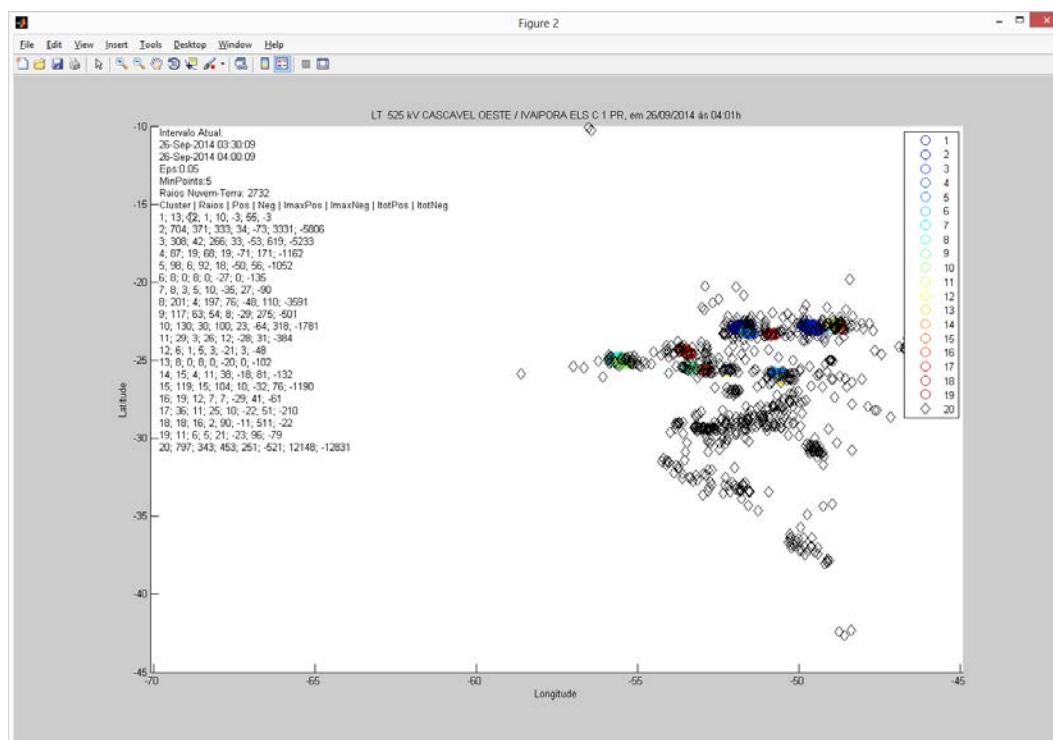


Figura 45 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 03:30:09 a 04:00:09)

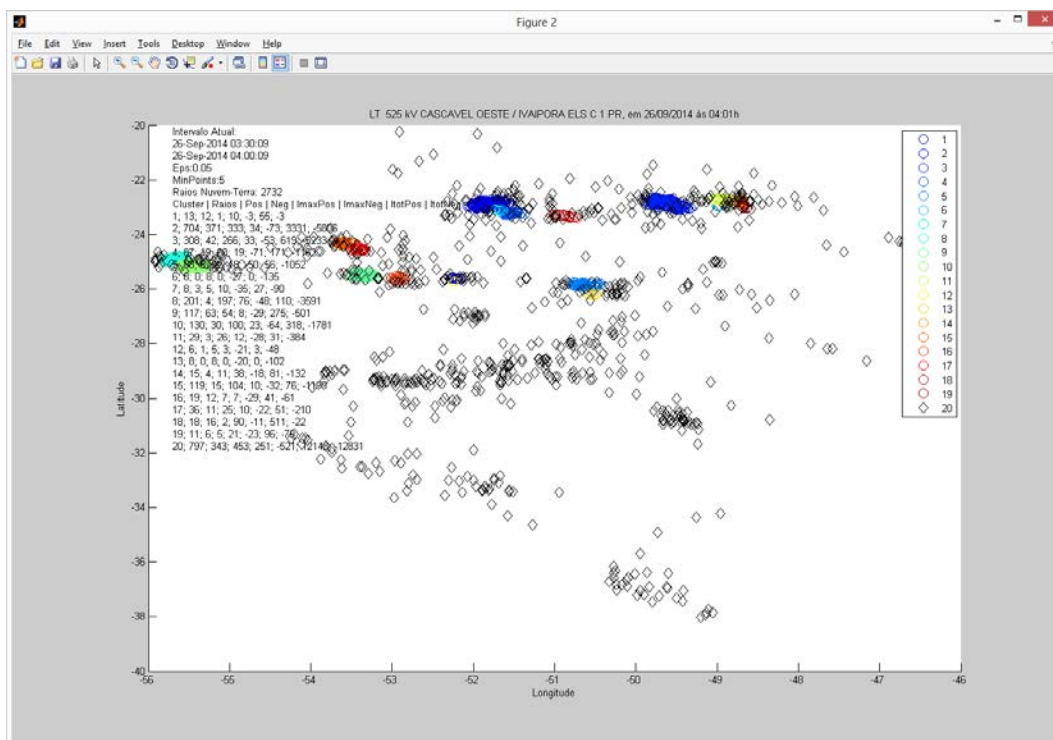


Figura 46 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 03:30:09 a 04:00:09)

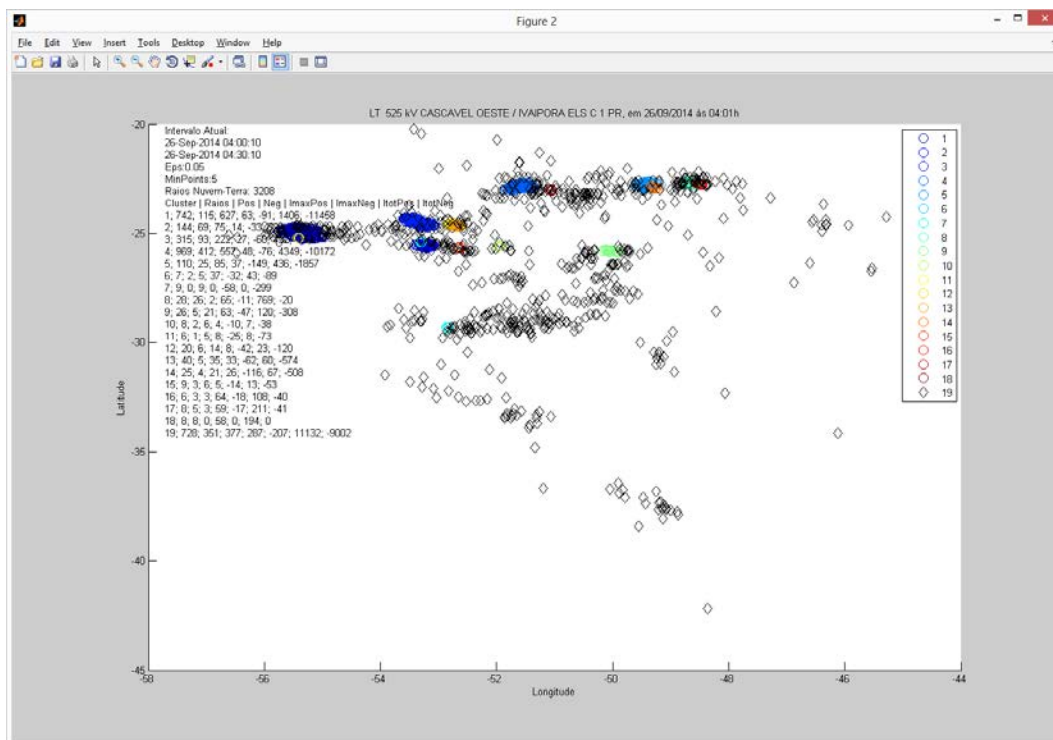


Figura 47 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 04:00:10 a 04:30:10)

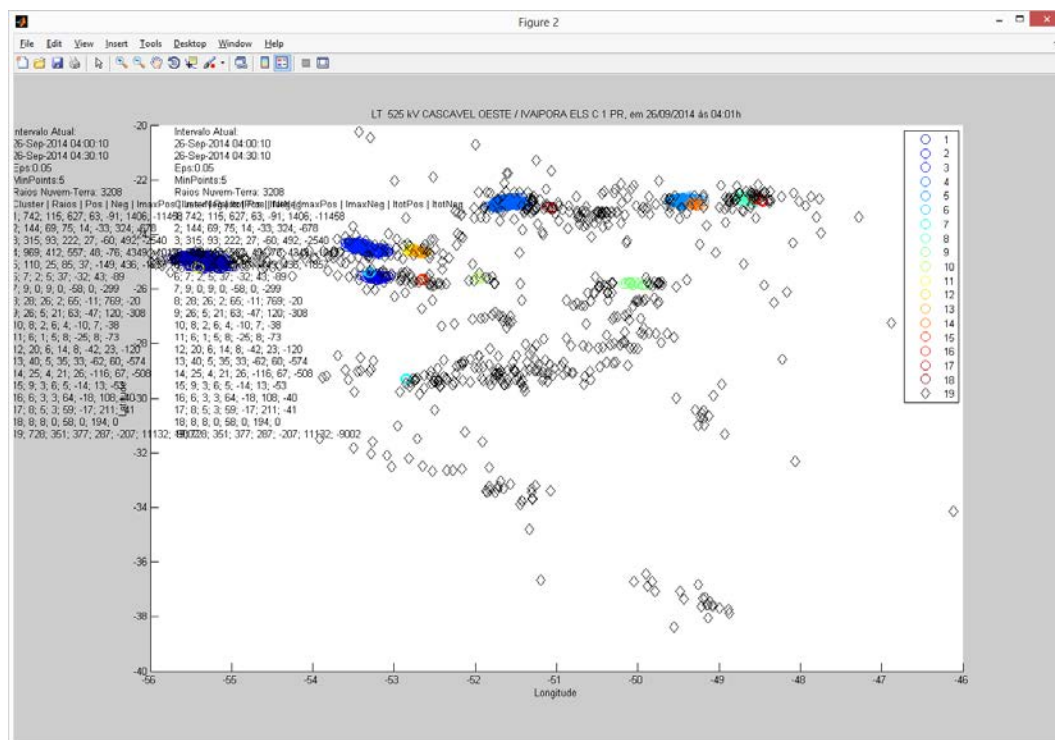


Figura 48 – DBSCAN para Eps=0,05 e MinPts=5 (Intervalo 04:00:10 a 04:30:10)

I. Agrupamento do espaço amostral de descargas atmosféricas - Estudo de Caso

As figuras abaixo representam as 31 cenas com os agrupamentos das descargas atmosféricas para os desligamentos à linha de transmissão LT 440 kV TAQUARUÇU / JUPIA, gerados através do método de agrupamento baseado em densidade (DBSCAN). Também estão representadas as 31 cenas com agrupamentos de descargas atmosféricas não associadas ao desligamento da referida linha de transmissão.

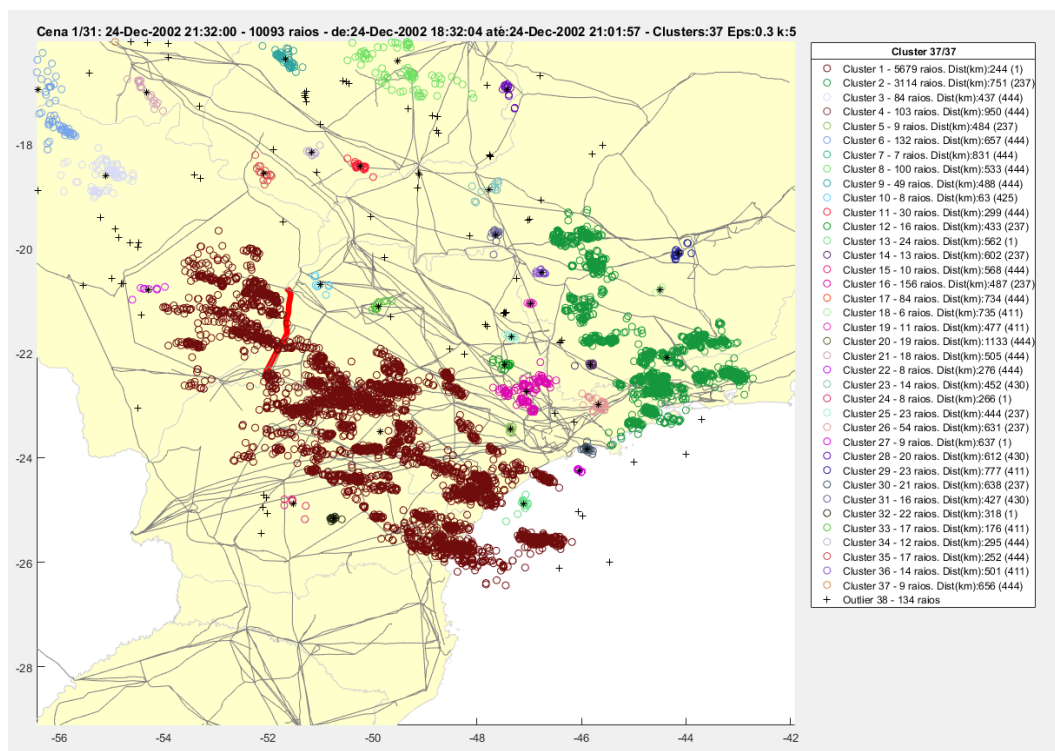


Figura 49 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 1

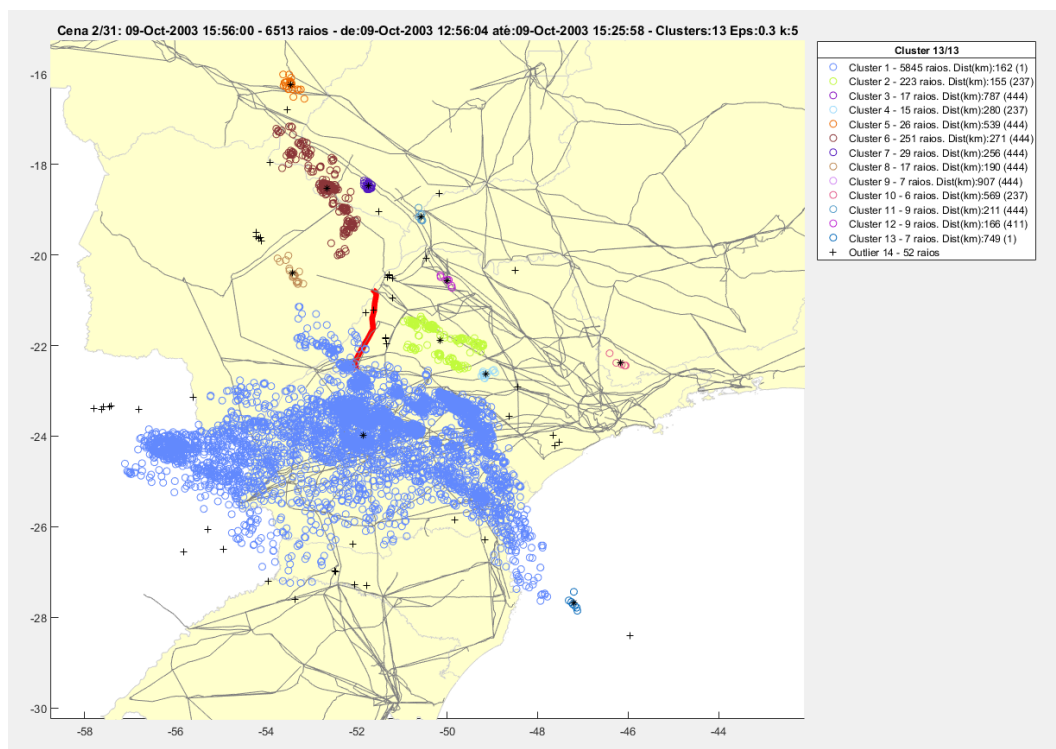


Figura 50 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 2

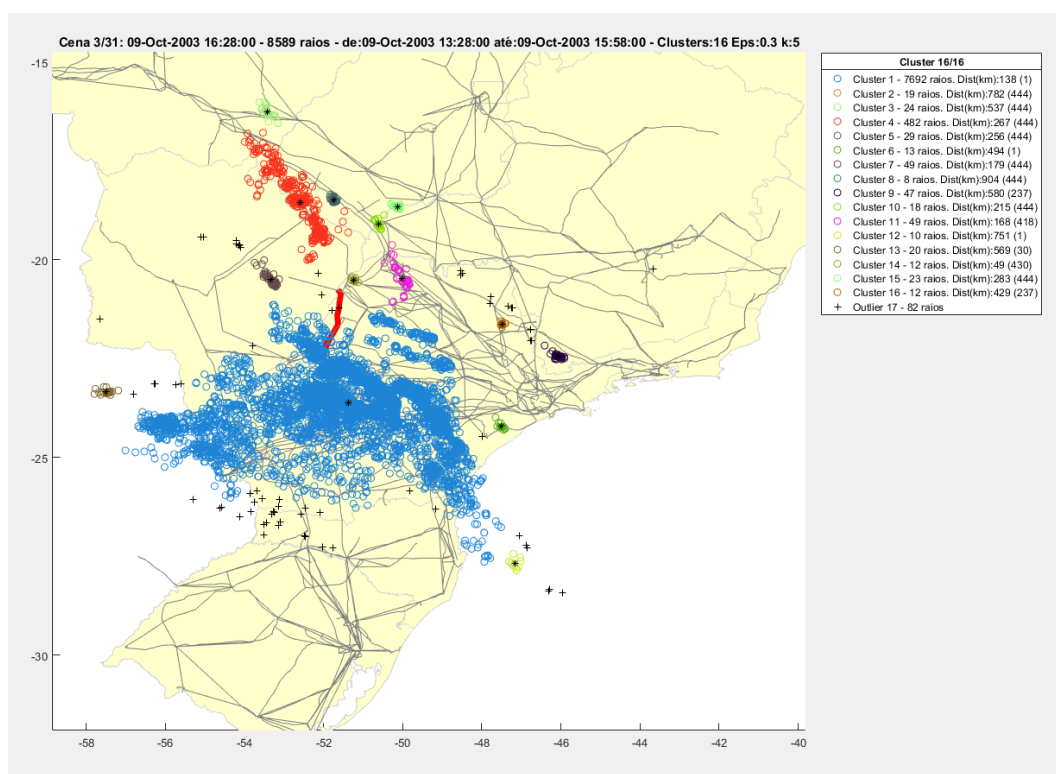


Figura 51 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 3

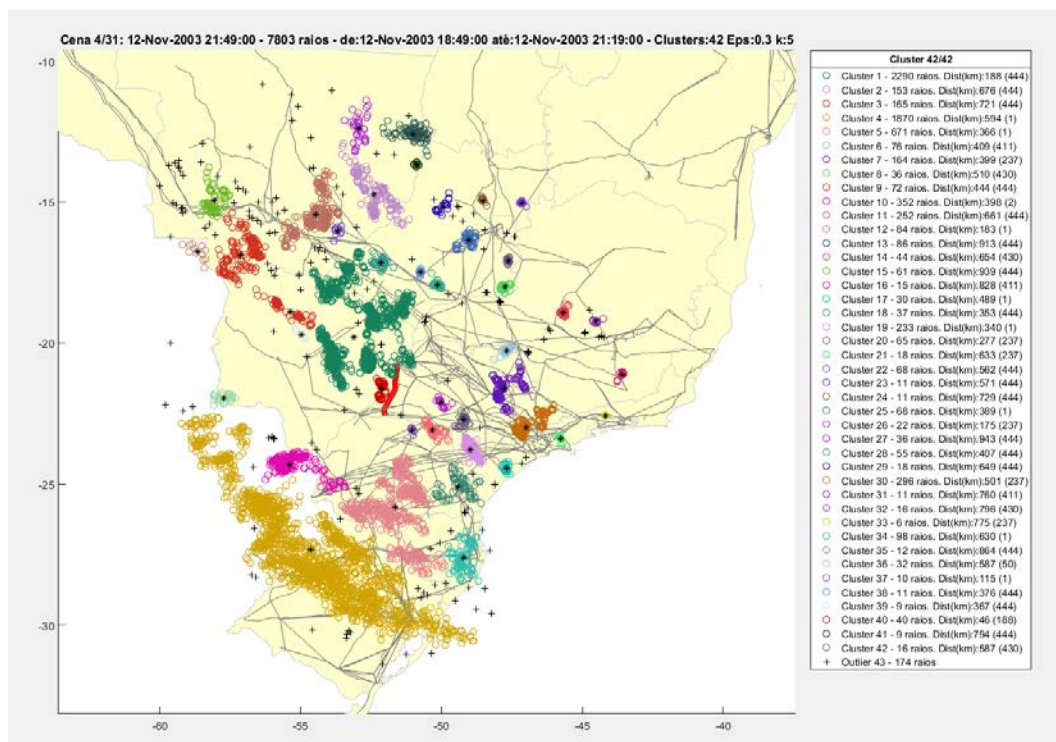


Figura 52 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 4

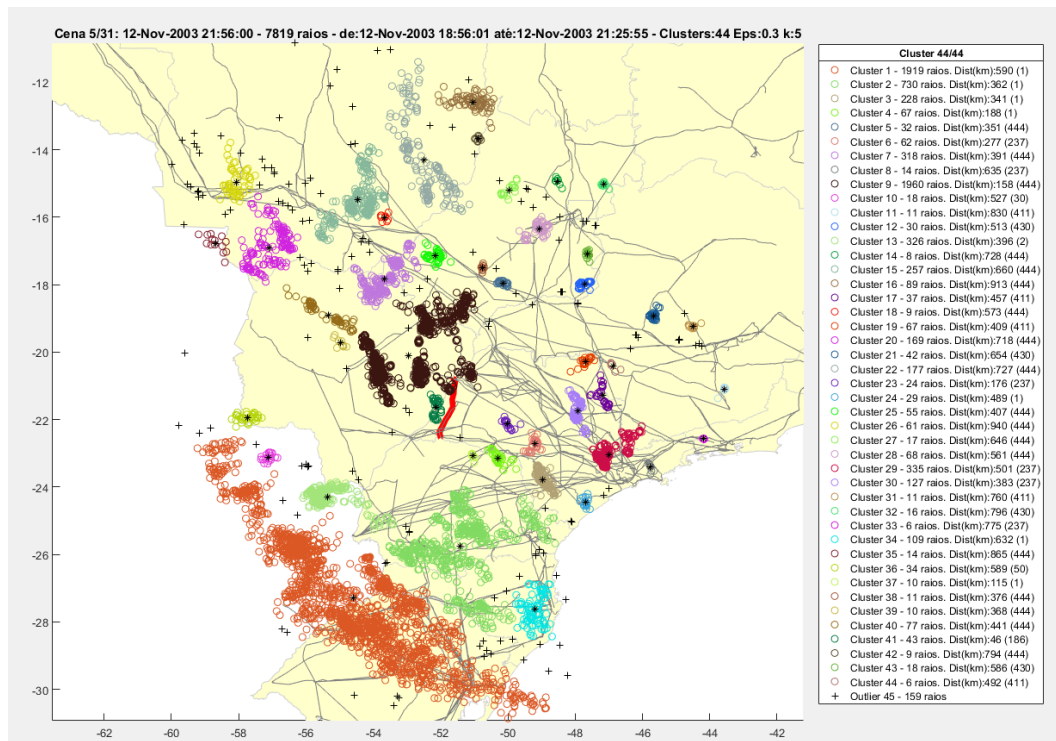


Figura 53 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 5

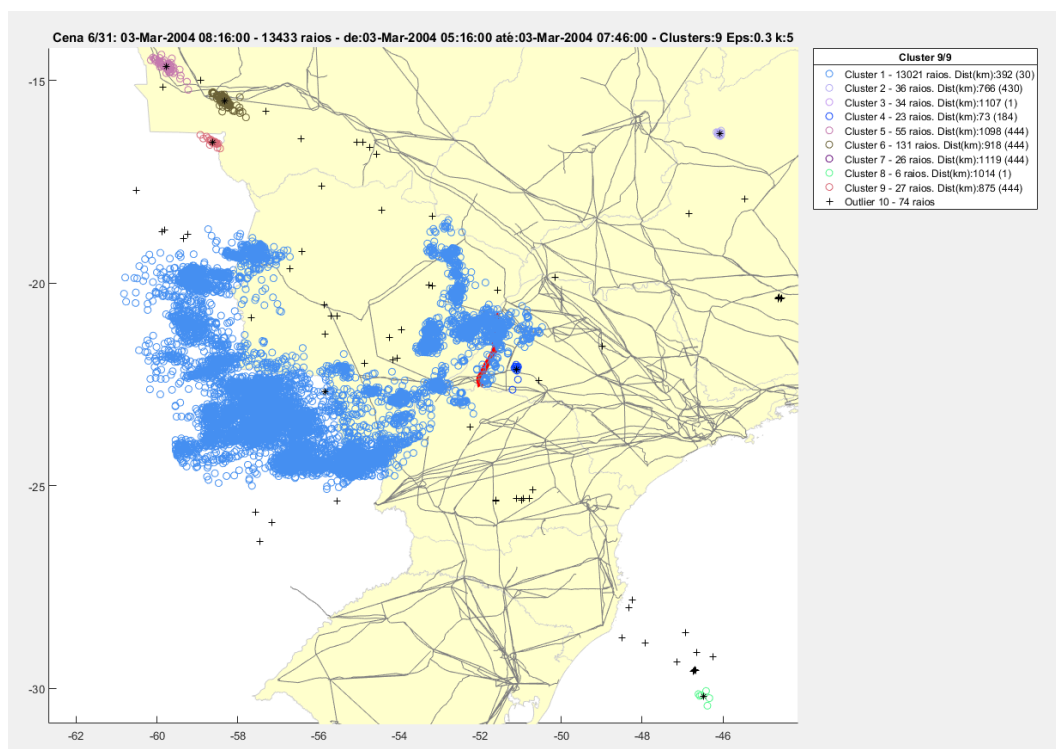


Figura 54 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 6

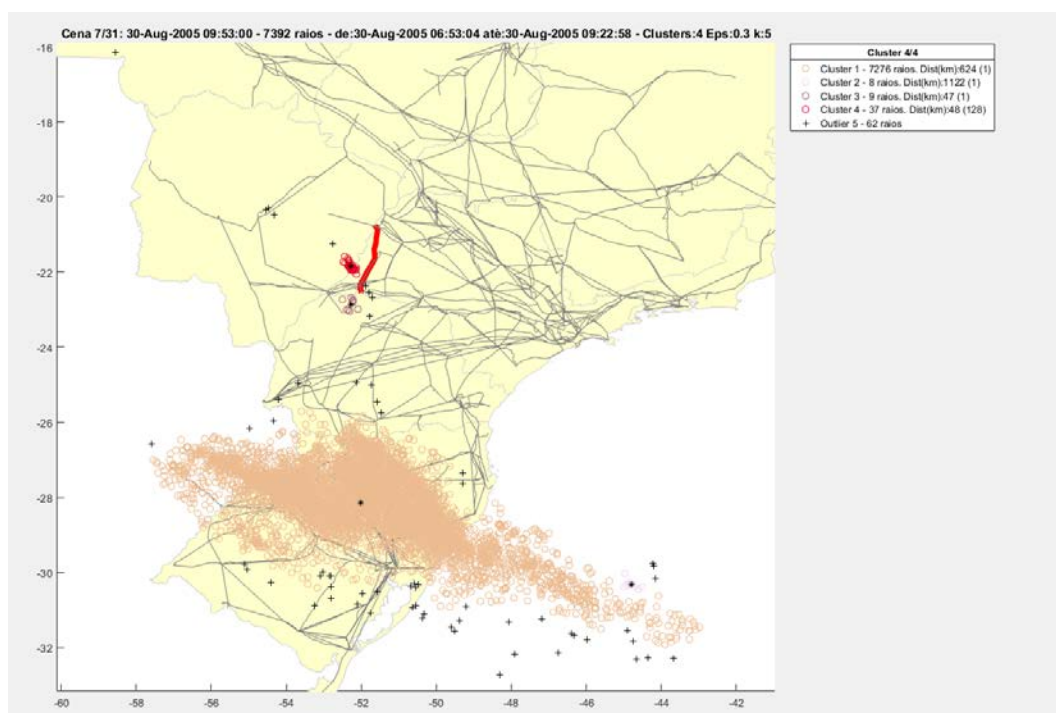


Figura 55 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 7

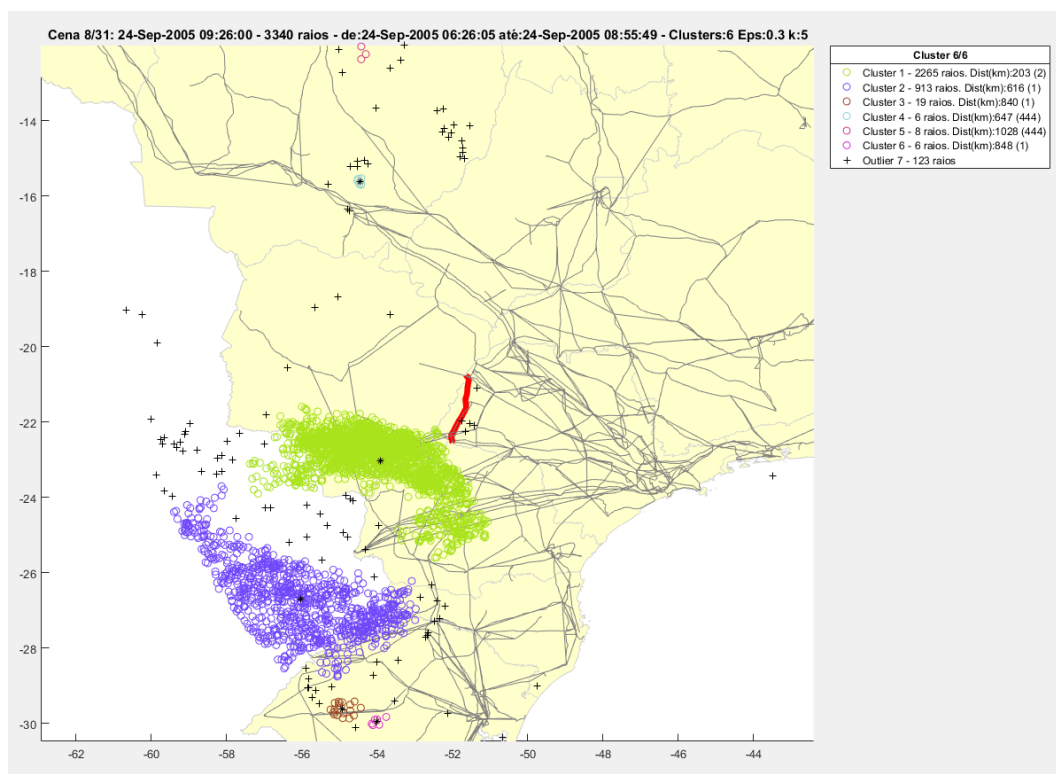


Figura 56 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 8

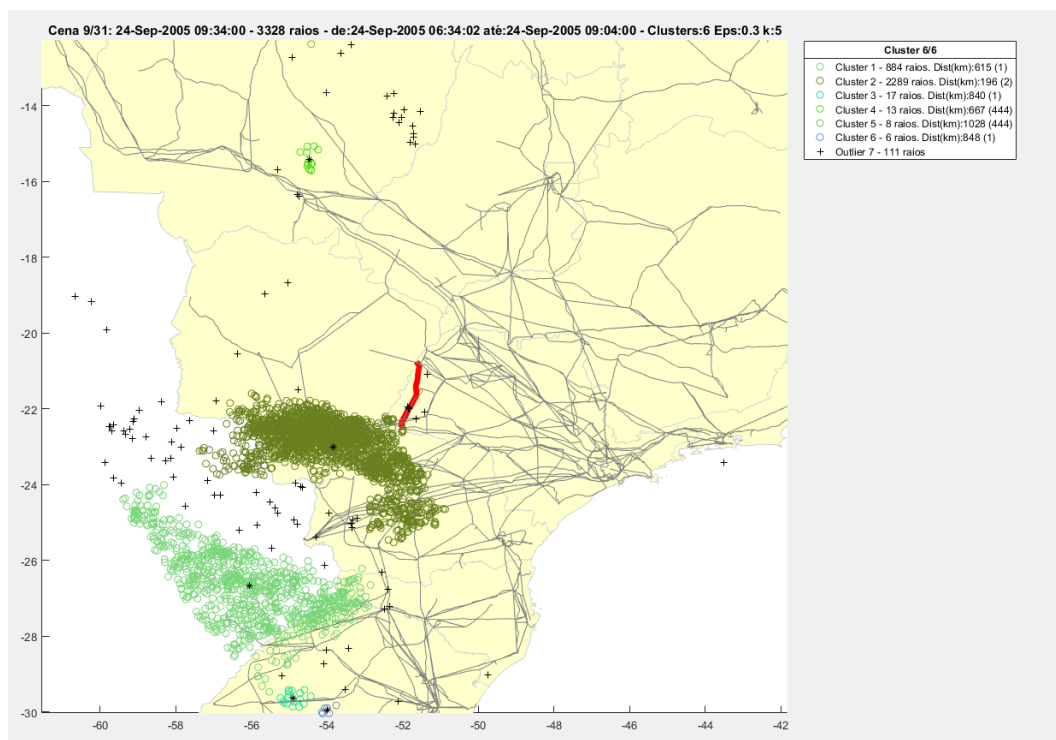


Figura 57 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 9

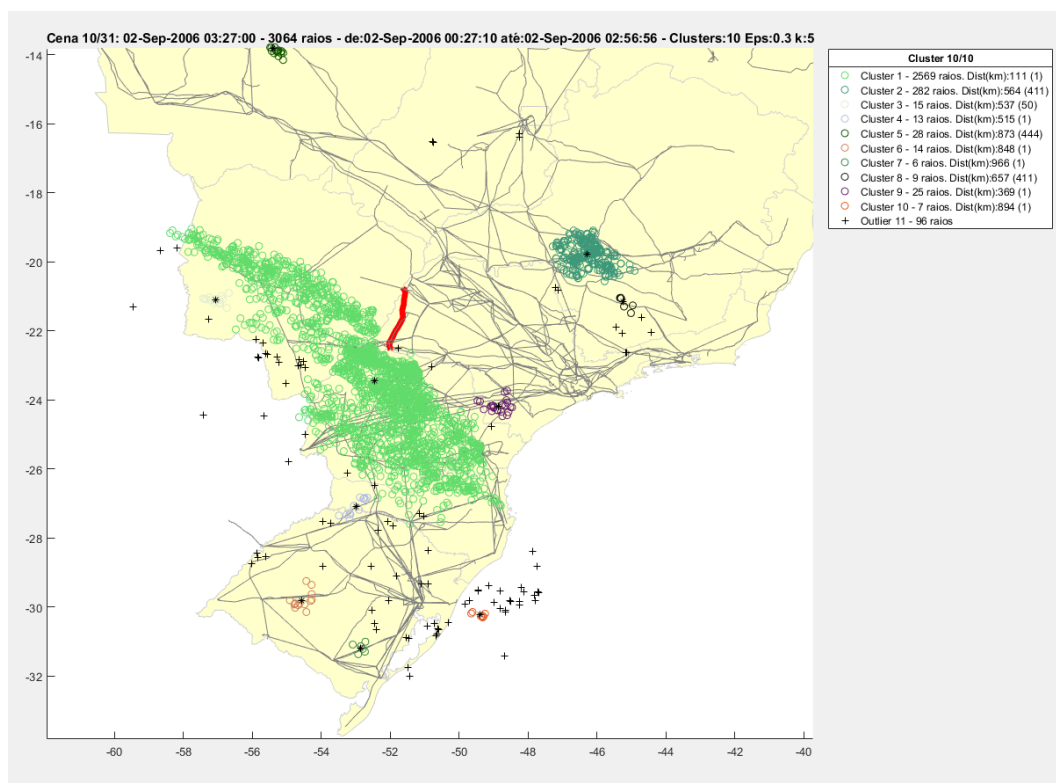


Figura 58 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 10

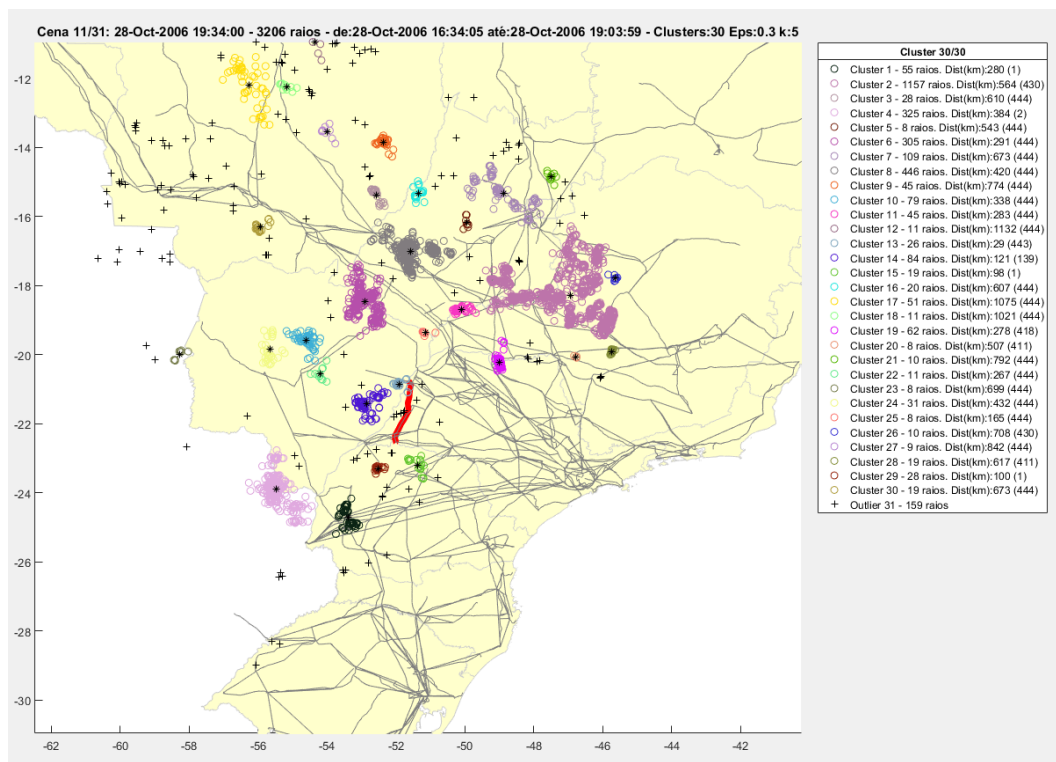


Figura 59 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 11

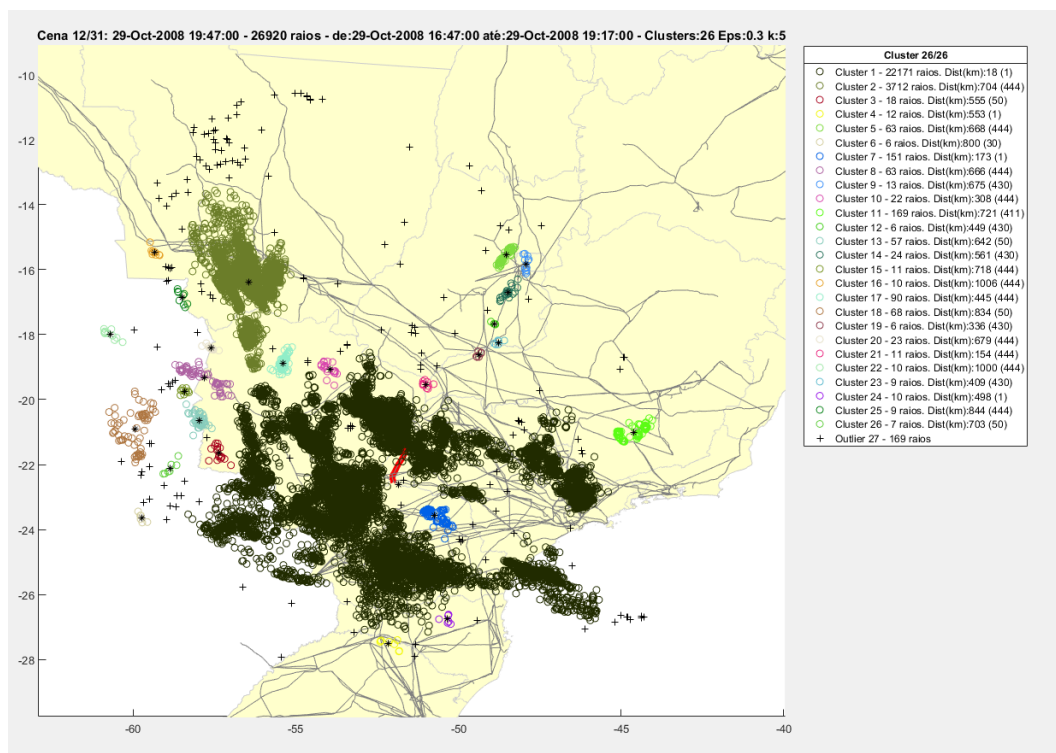


Figura 60 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 12

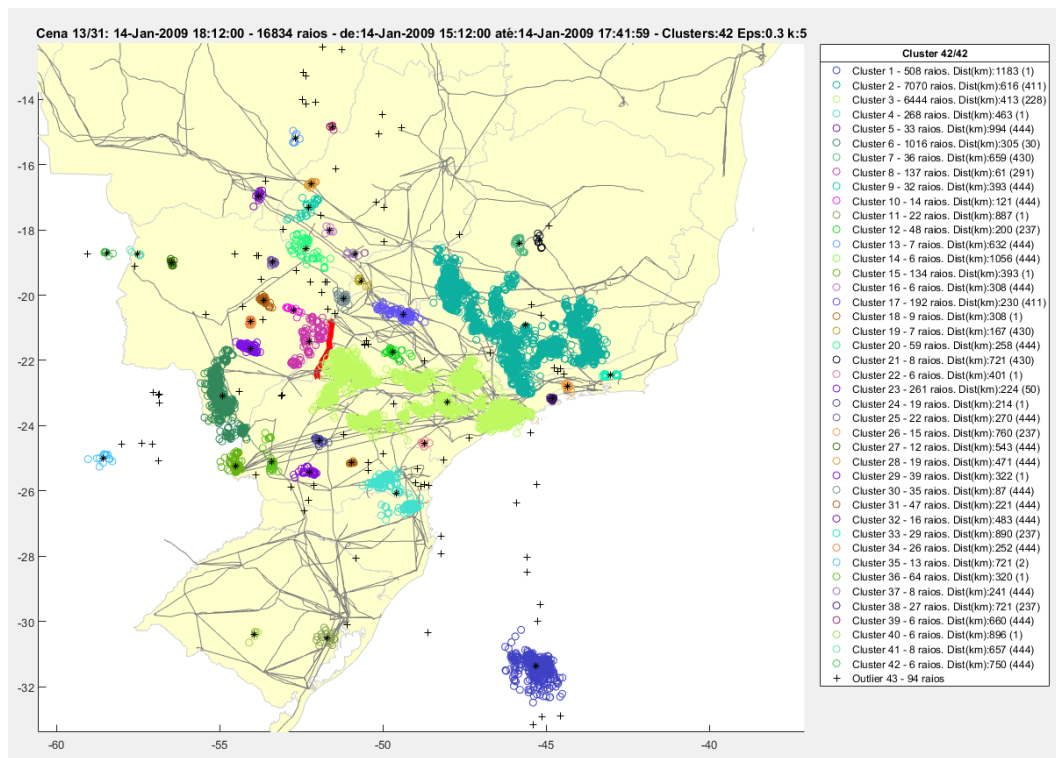


Figura 61 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 13

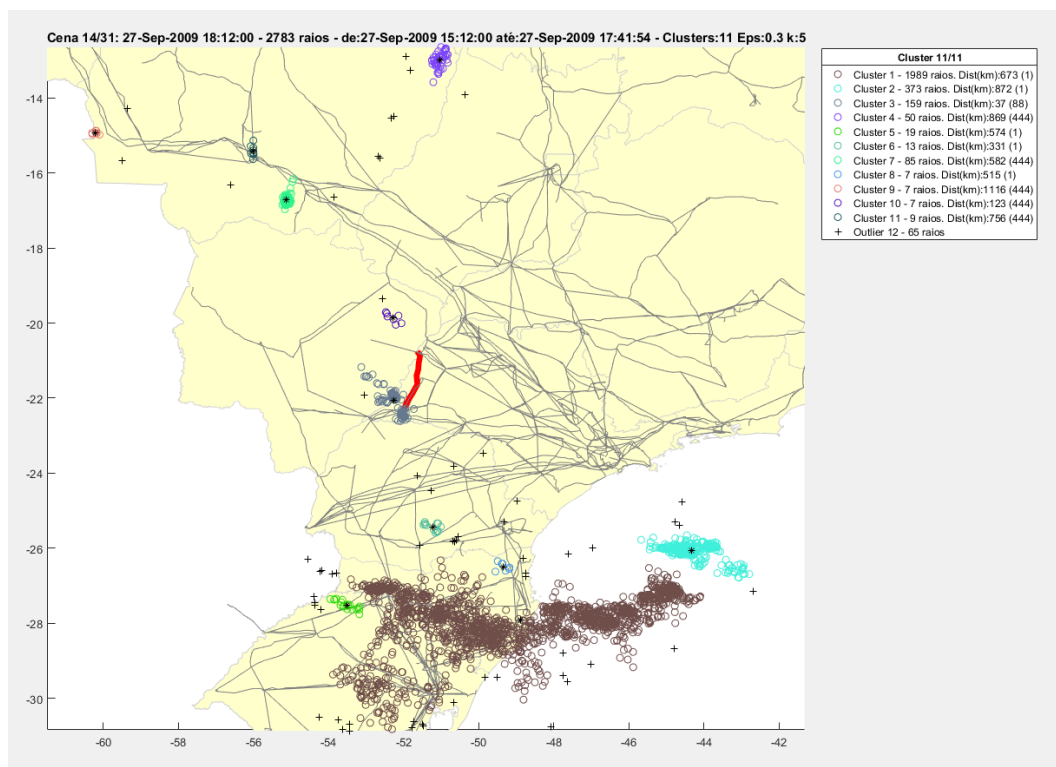


Figura 62 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 14

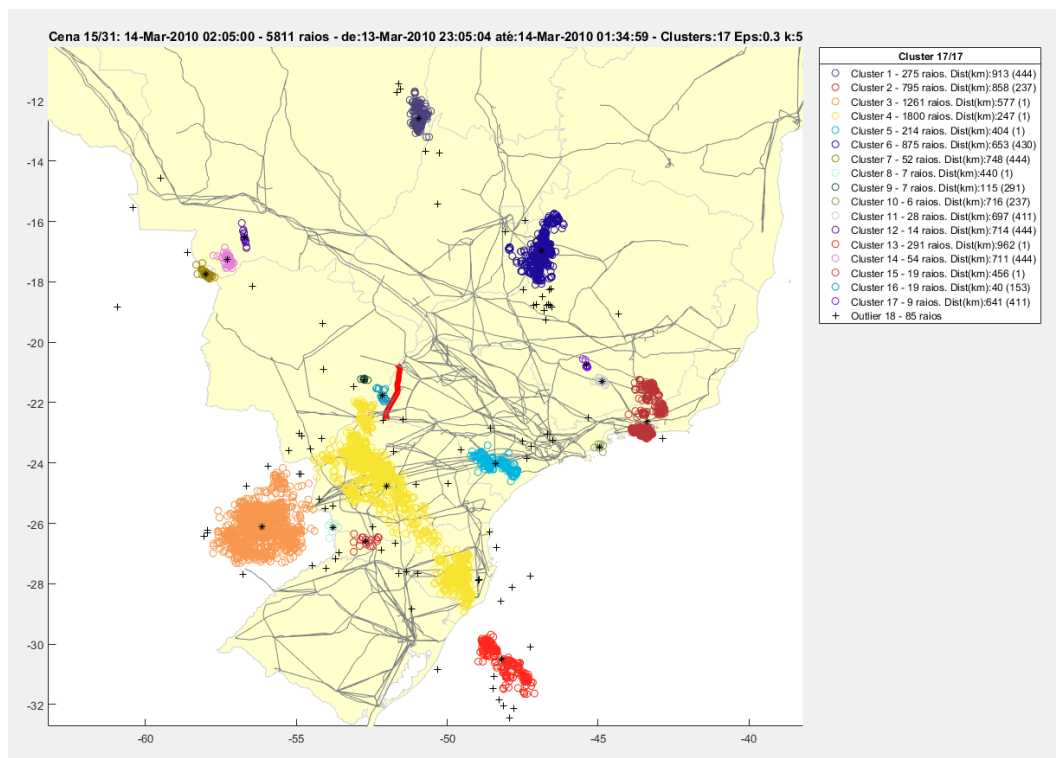


Figura 63 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 15

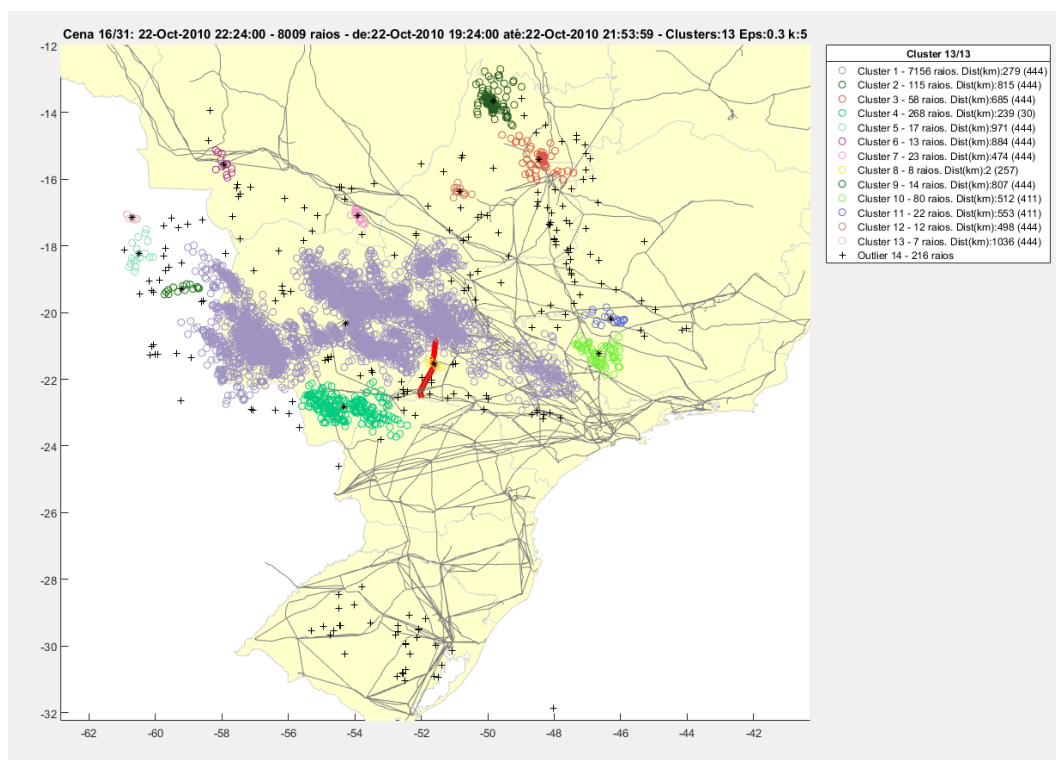


Figura 64 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 16

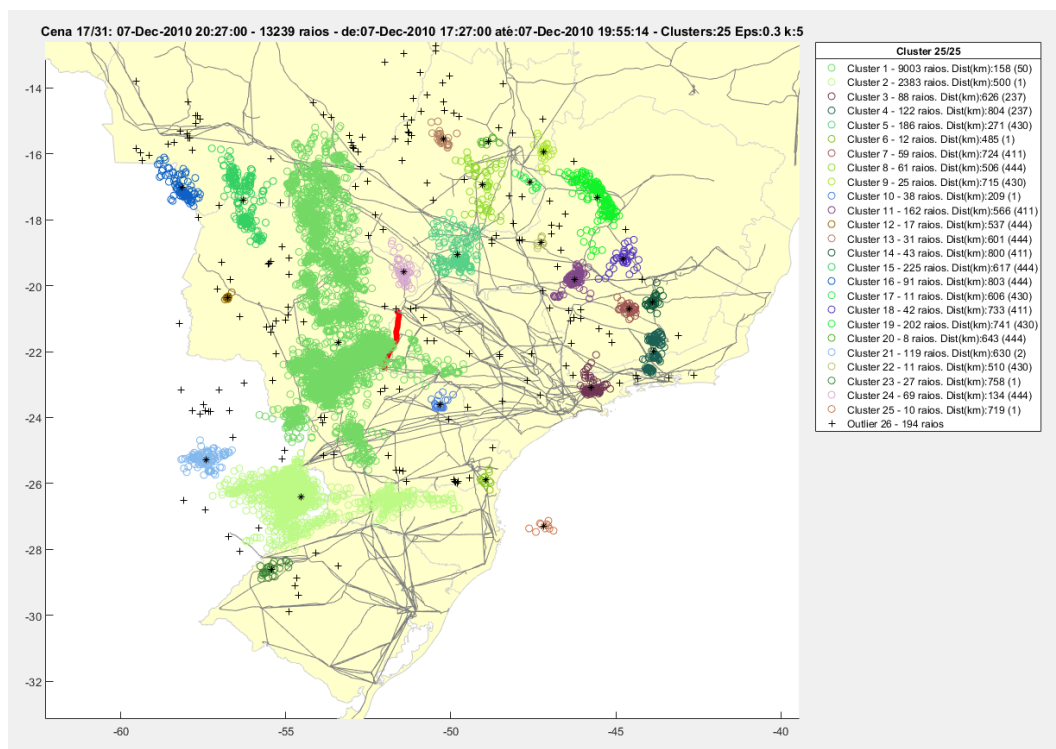


Figura 65 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 17

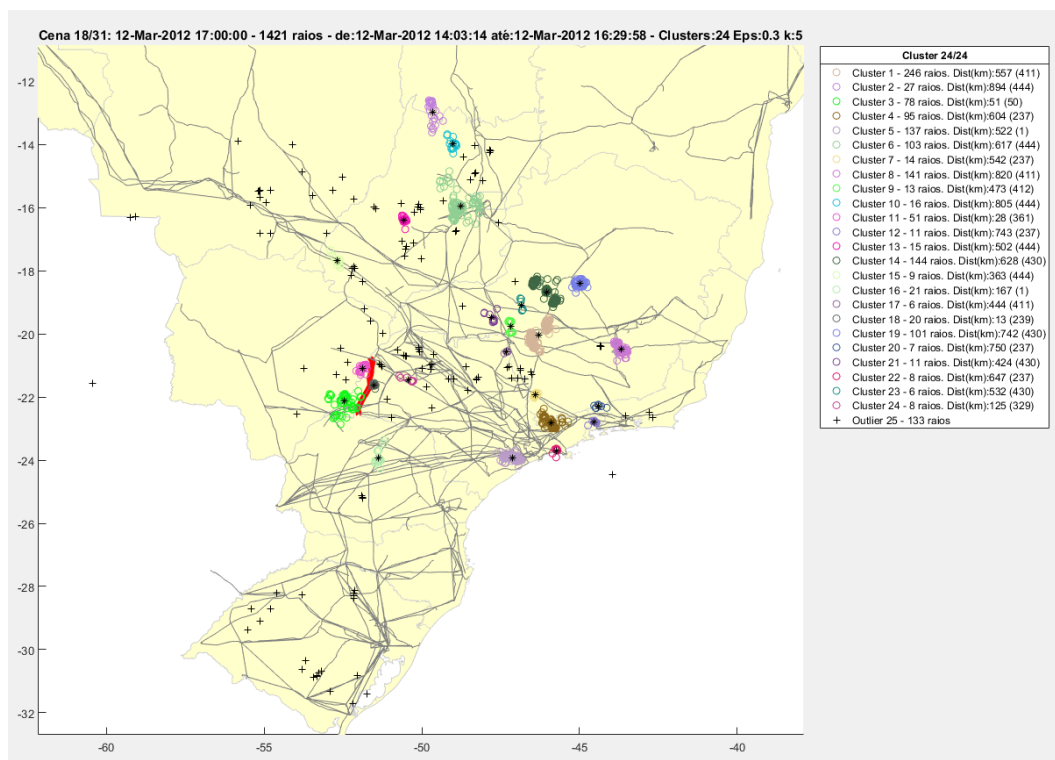


Figura 66 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 18

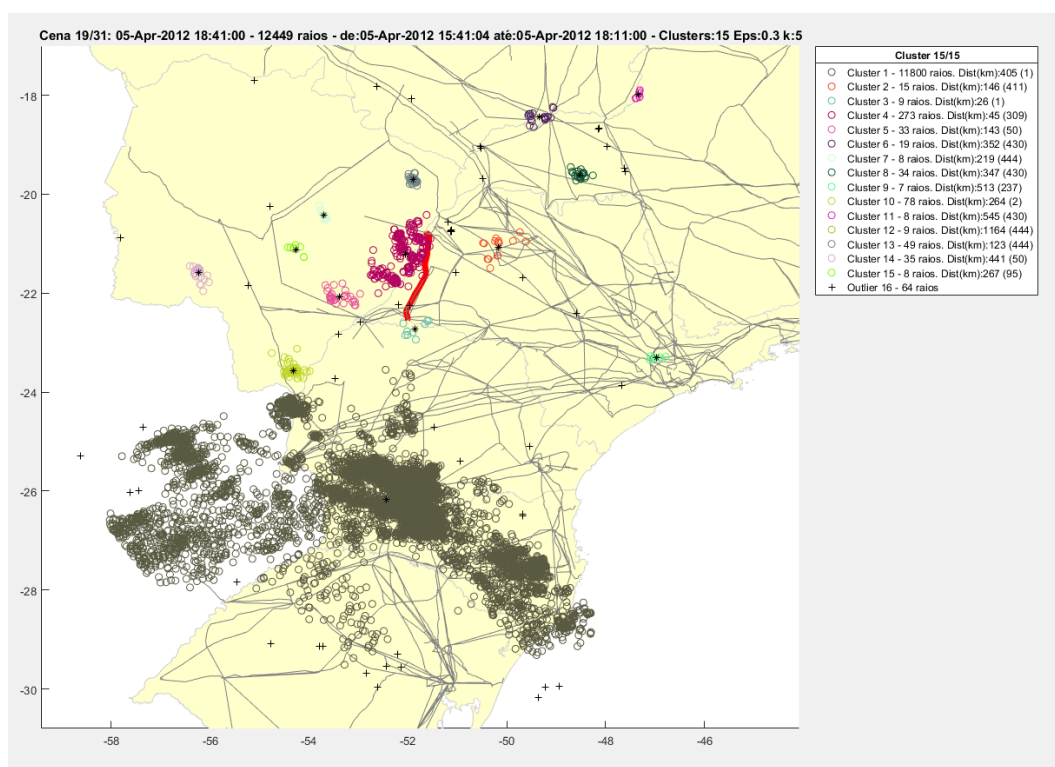


Figura 67 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 19

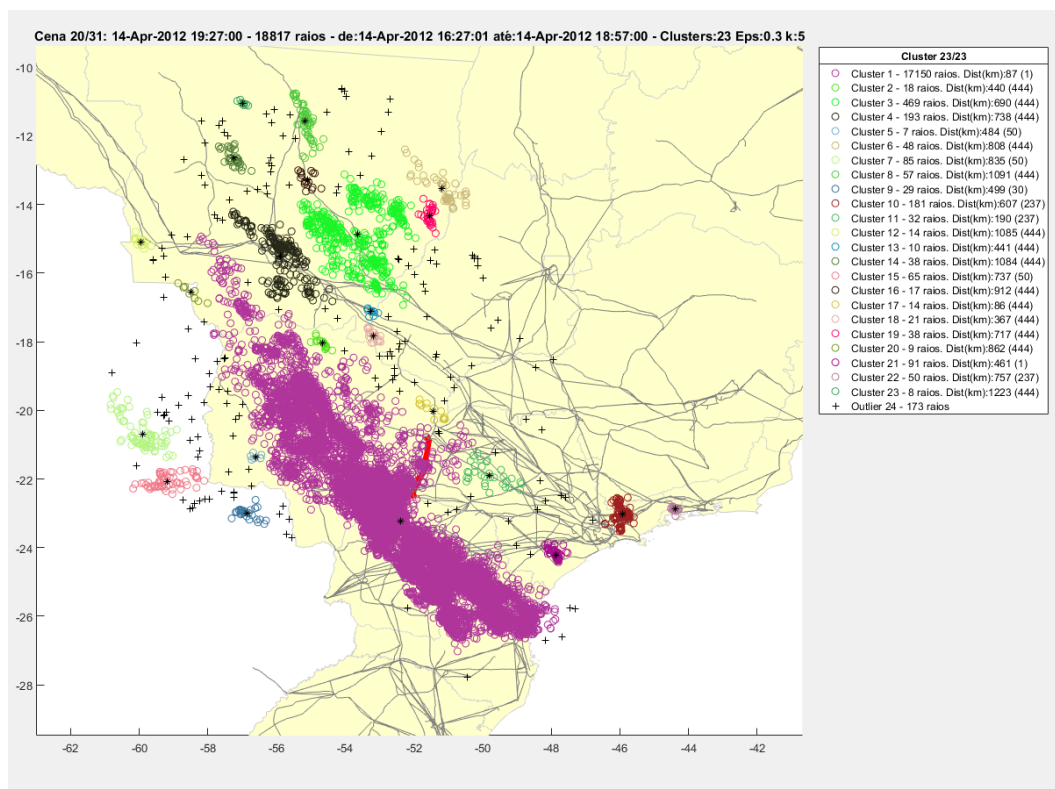


Figura 68 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 20

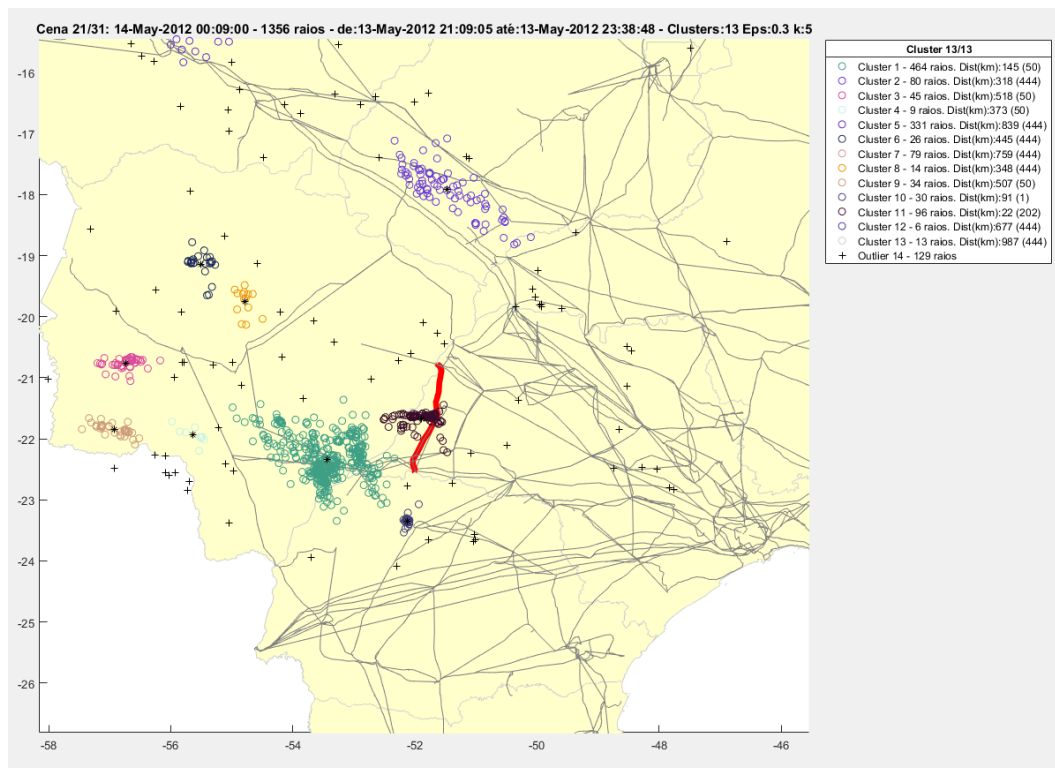


Figura 69 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 21

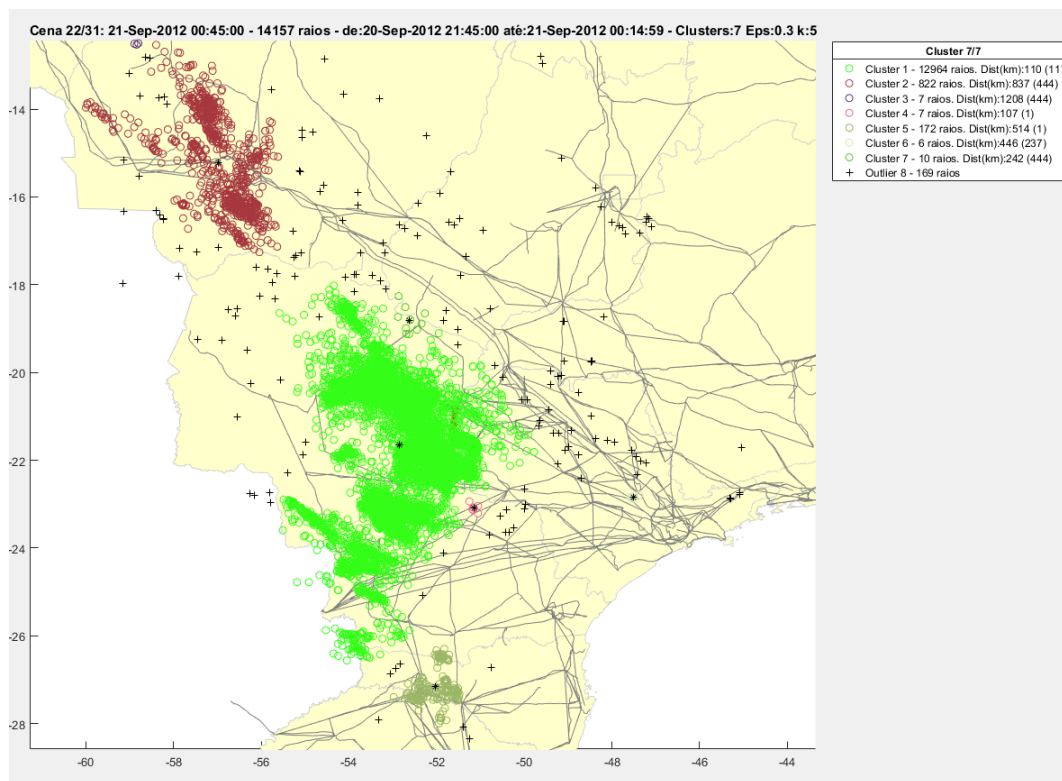


Figura 70 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 22

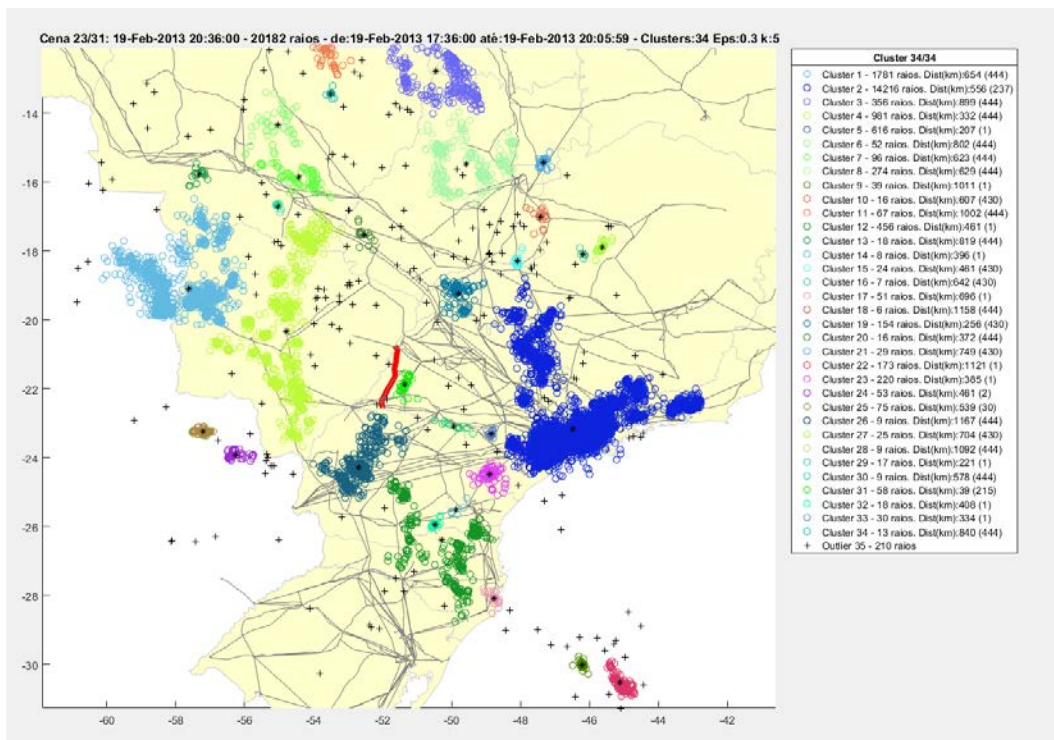


Figura 71 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 23

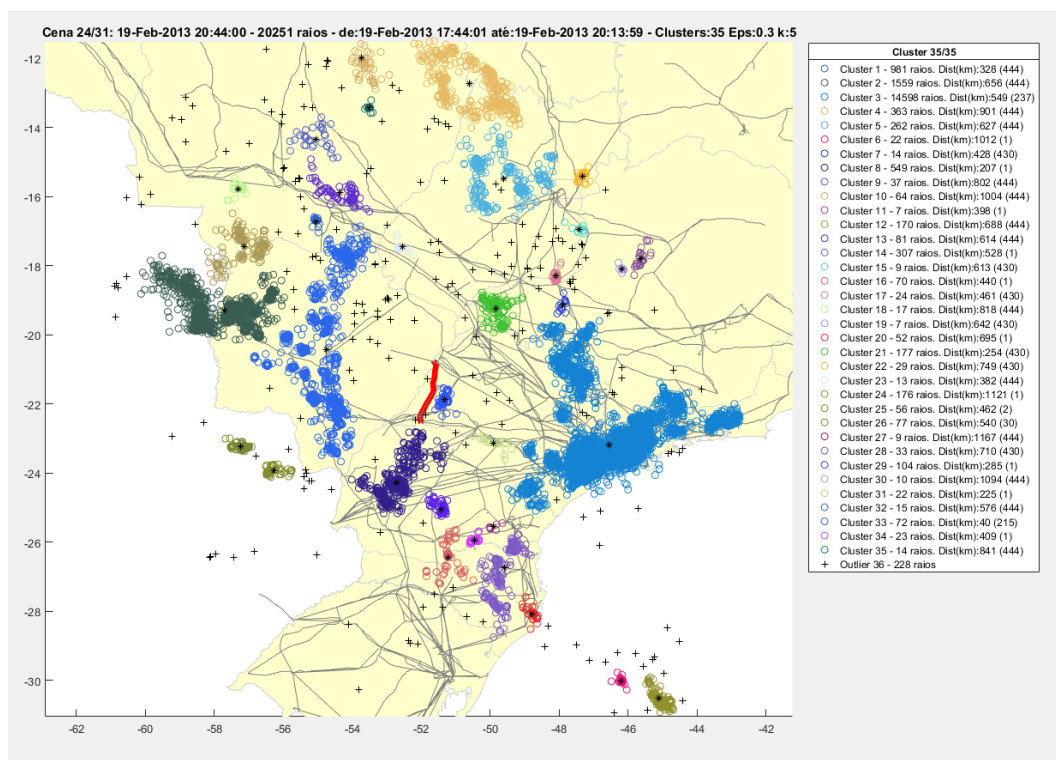


Figura 72 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 24

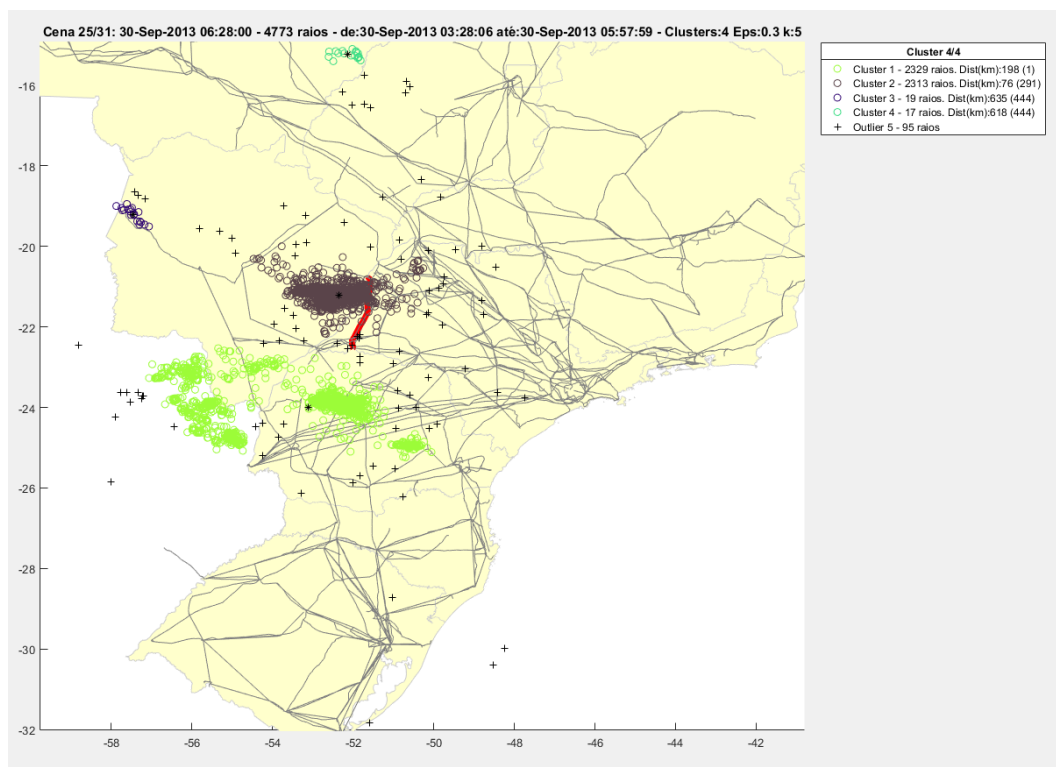


Figura 73 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 25

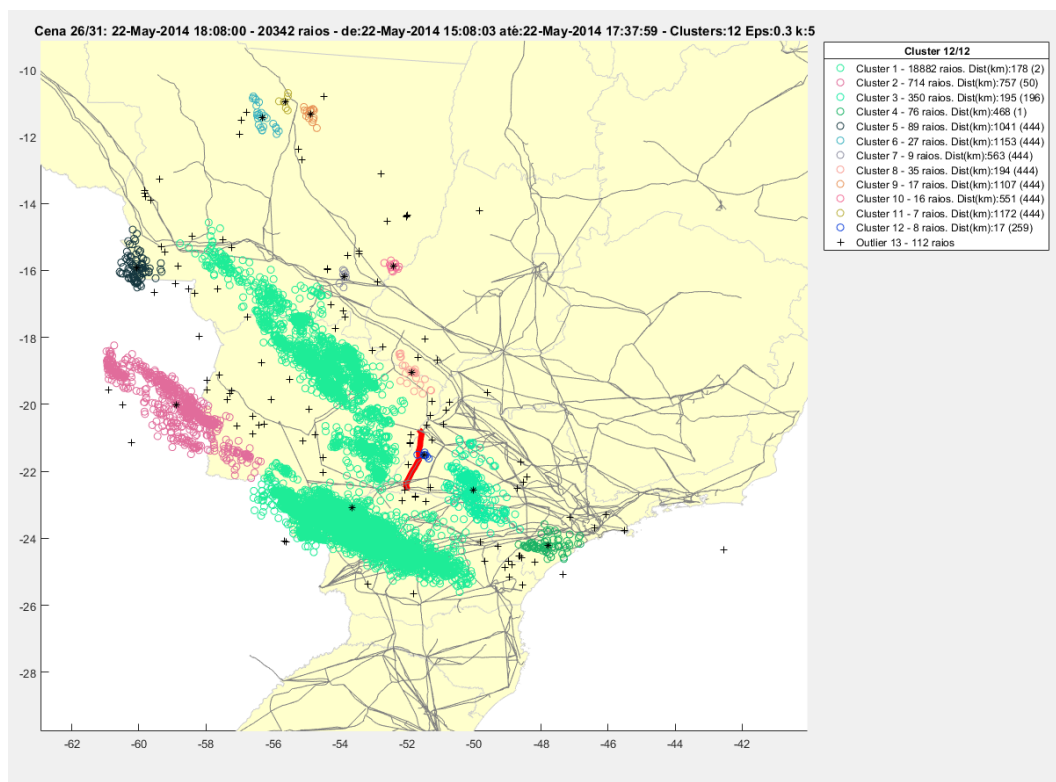


Figura 74 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 26

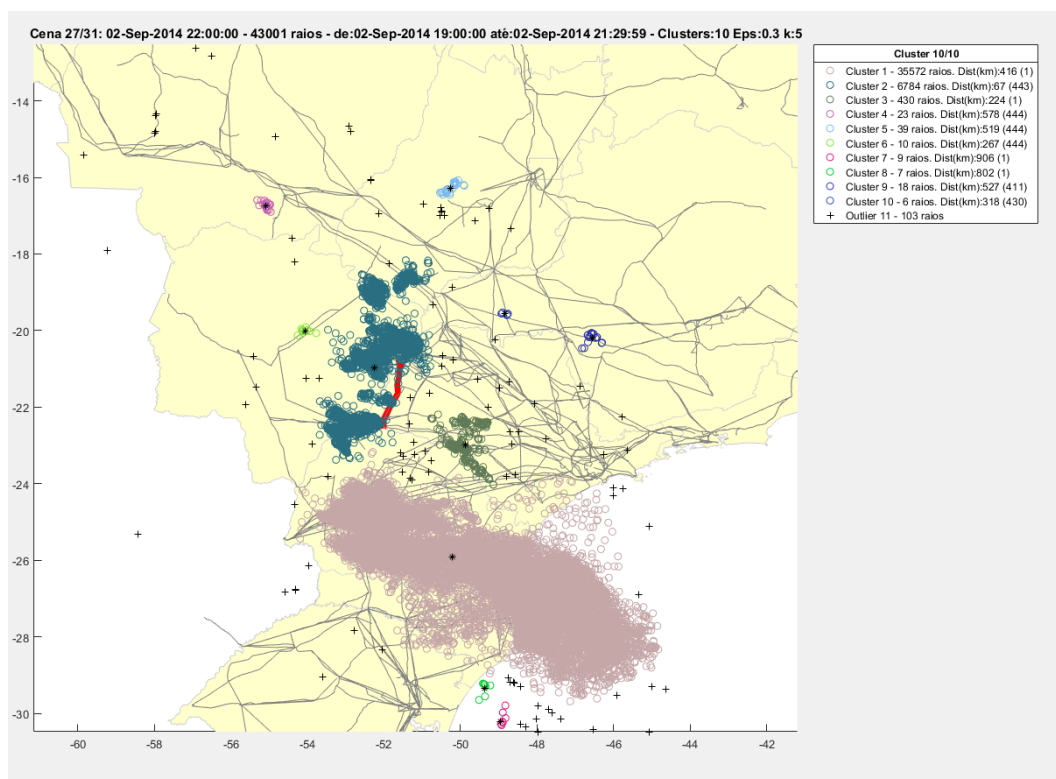


Figura 75 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 27

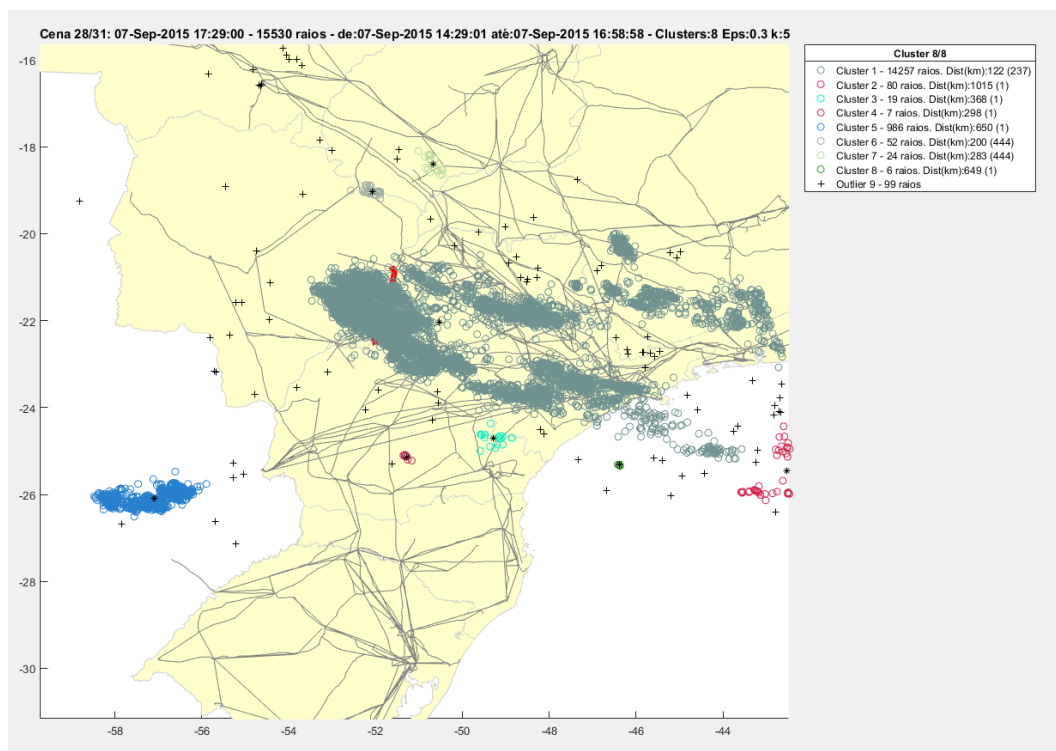


Figura 76 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 28

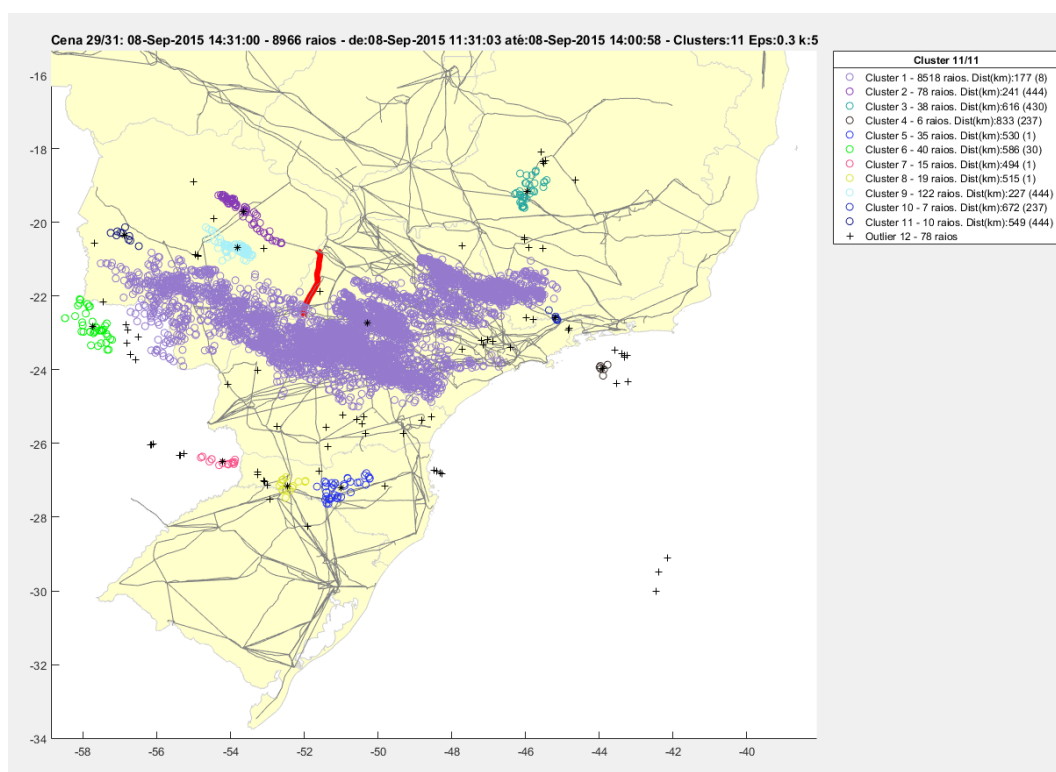


Figura 77 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 29

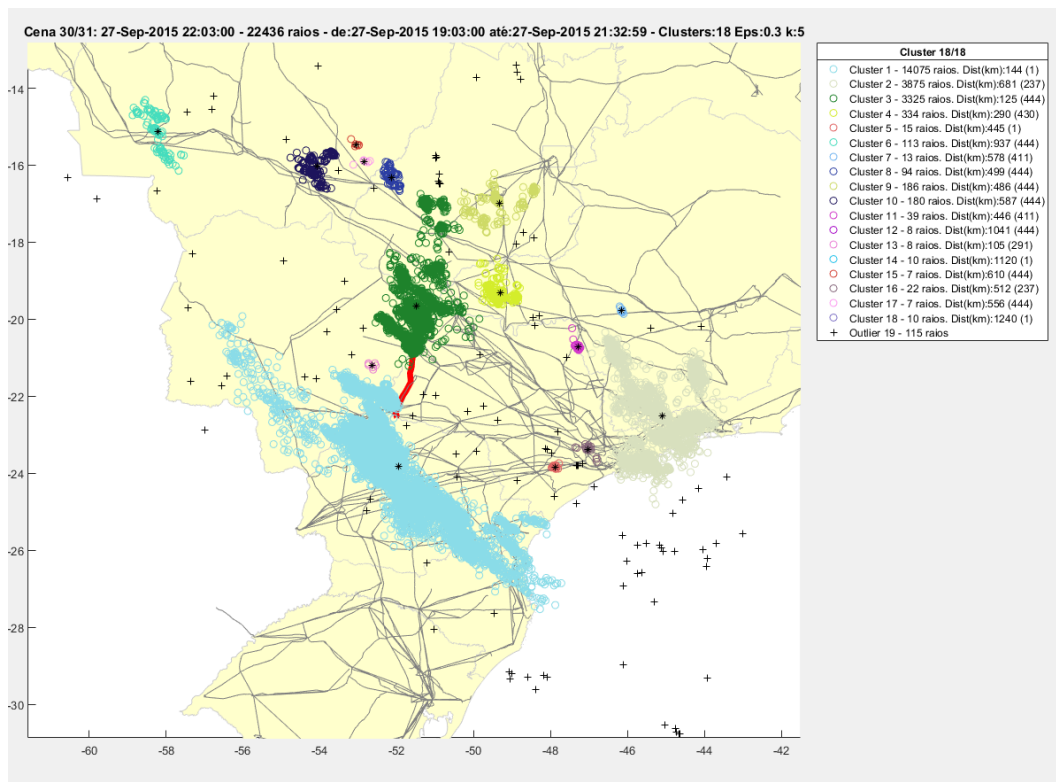


Figura 78 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 30

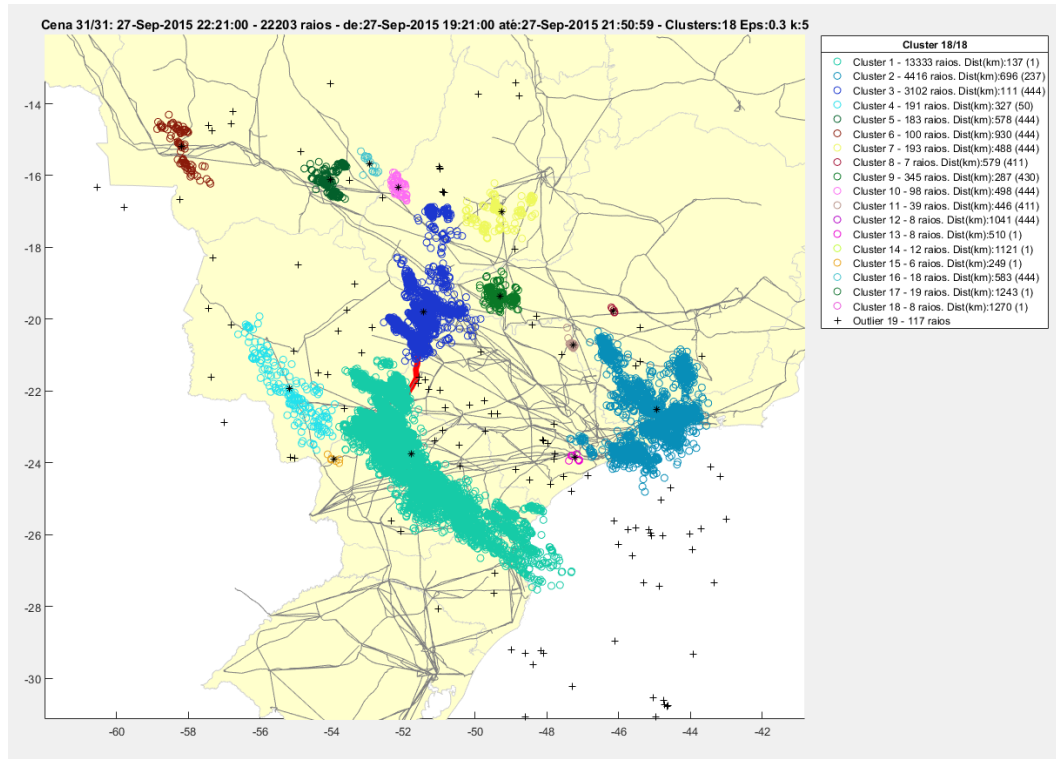


Figura 79 - Agrupamento das descargas atmosféricas associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 31

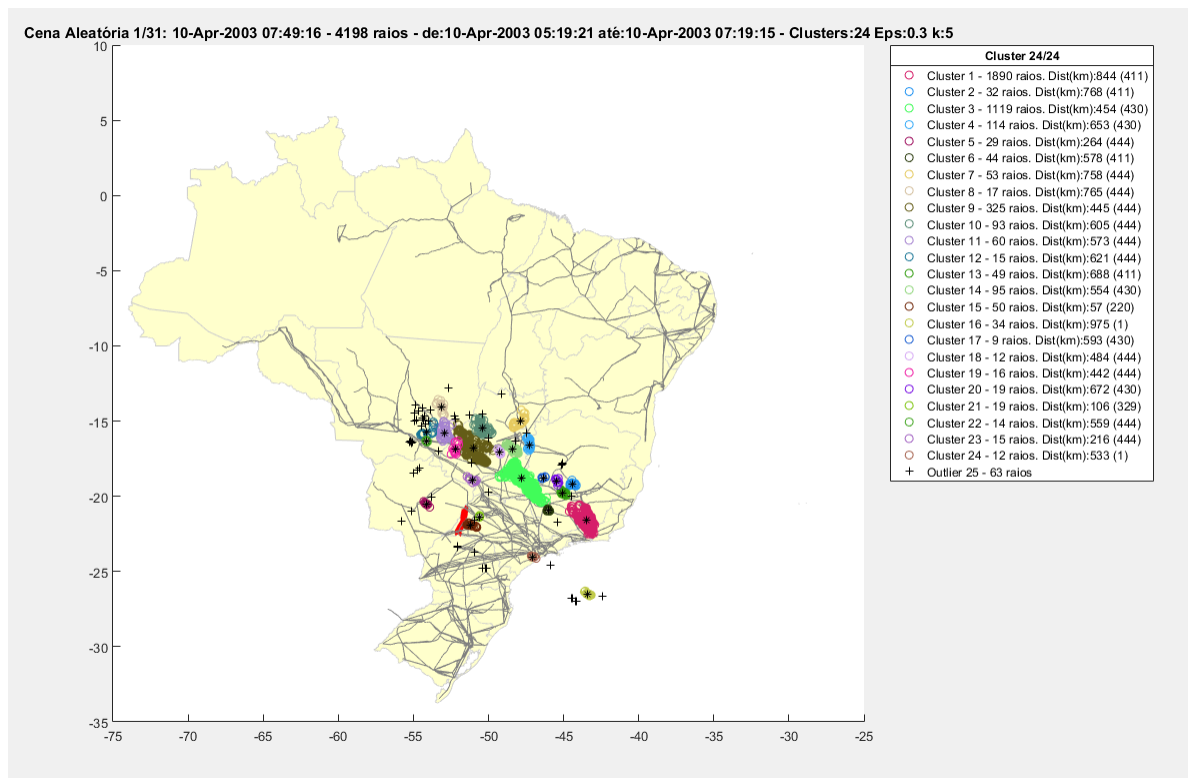


Figura 80 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 1

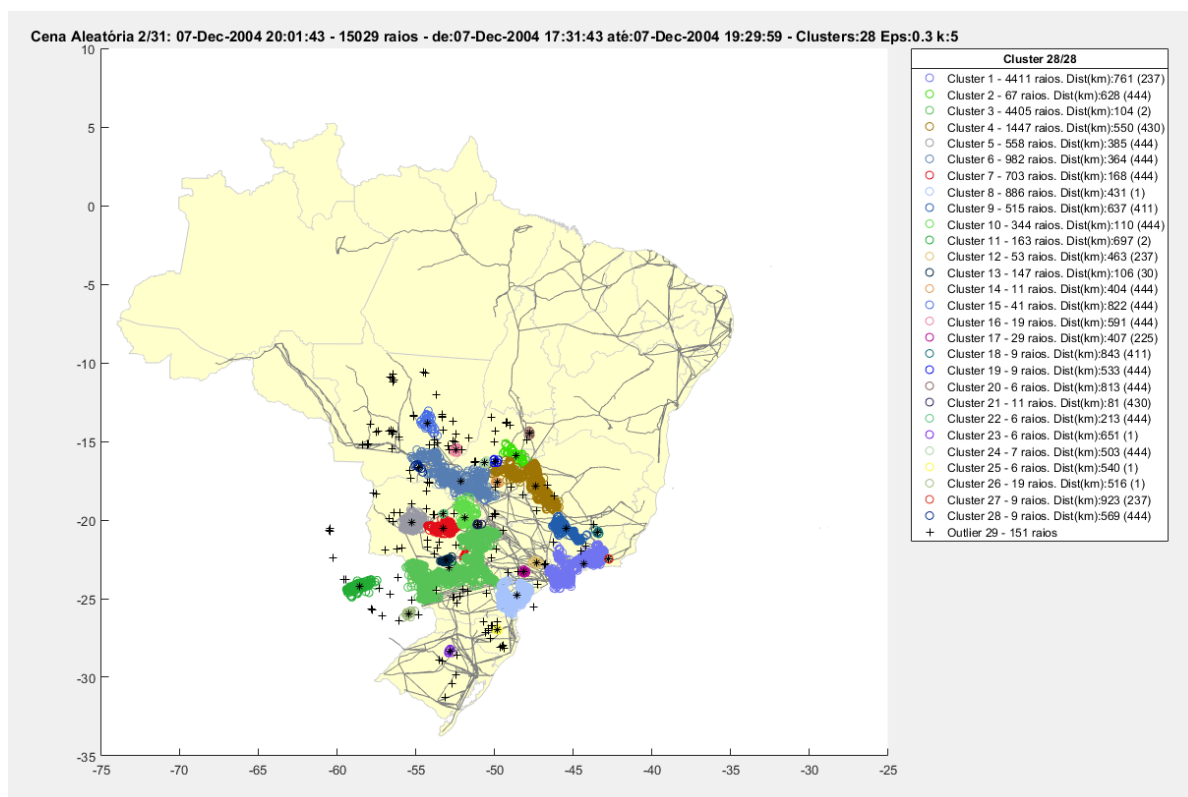


Figura 81 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 2

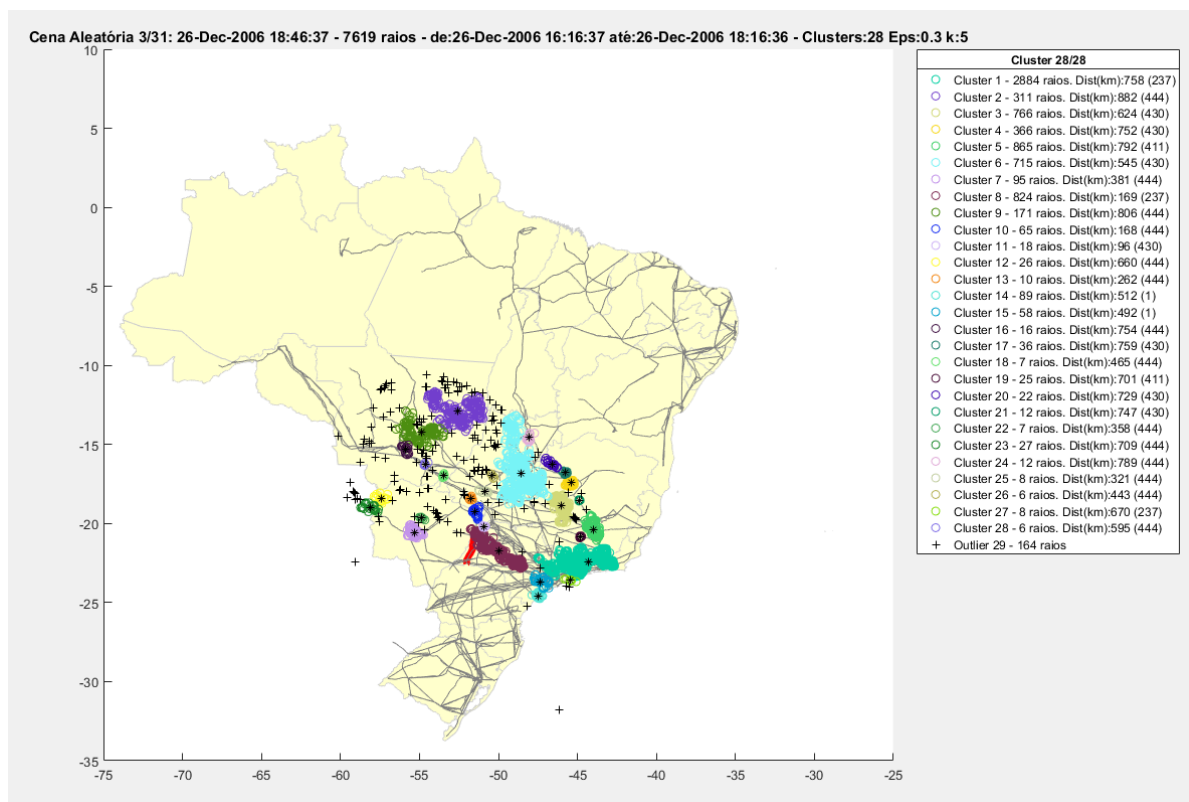


Figura 82 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 3

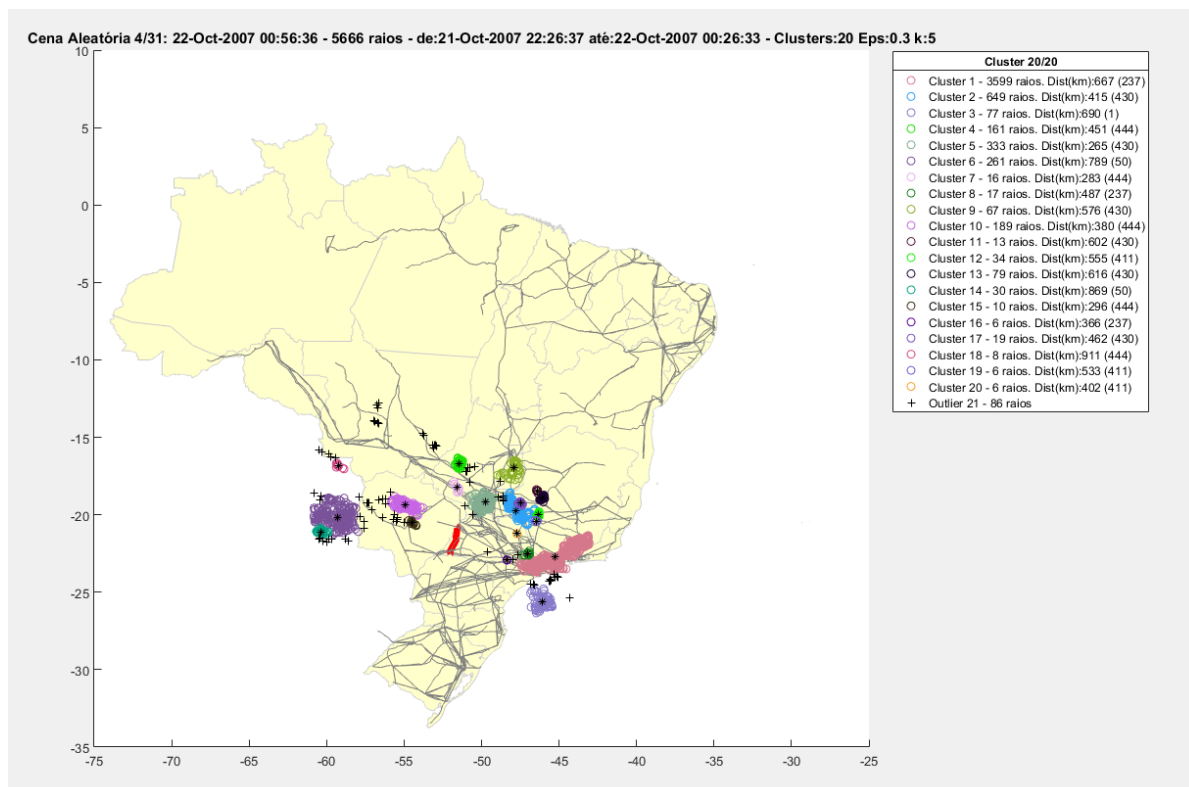


Figura 83 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 4

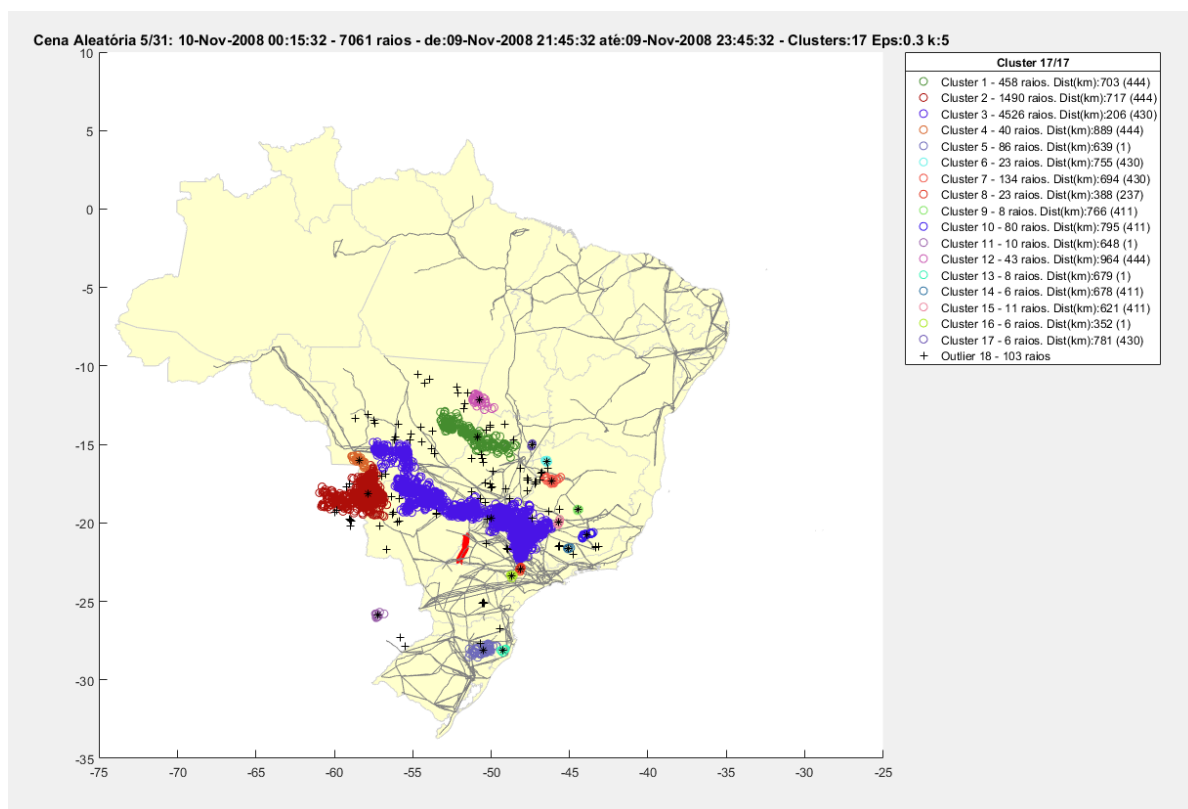


Figura 84 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 5

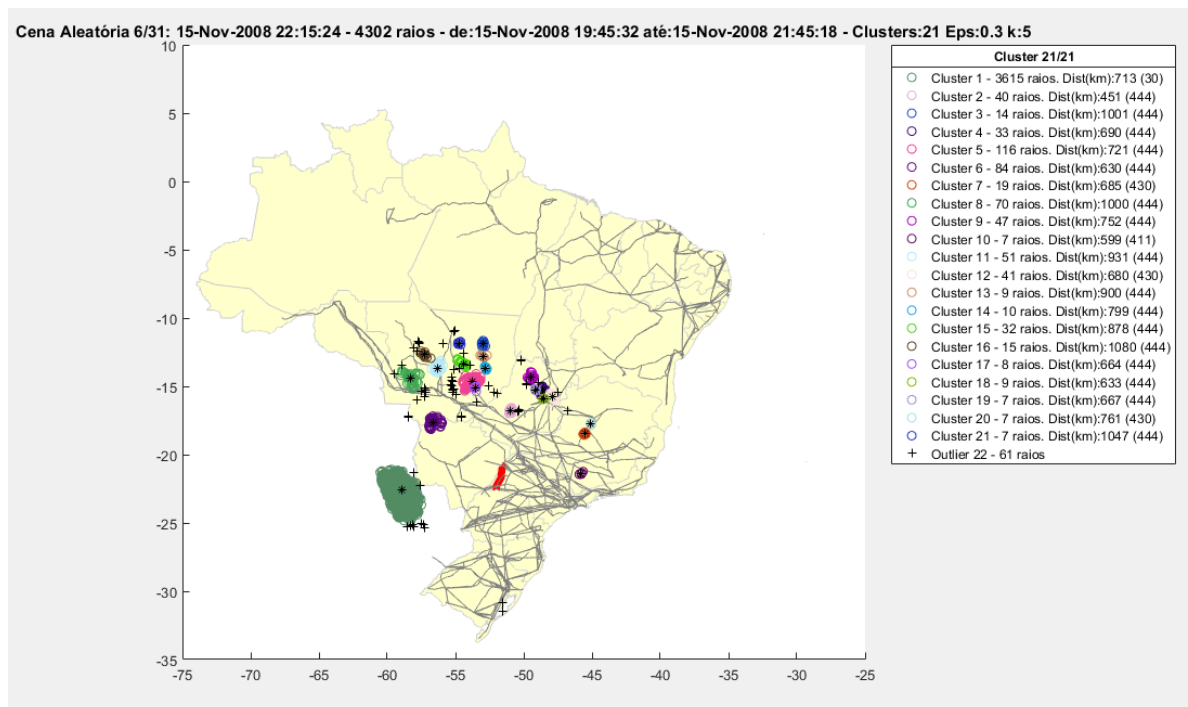


Figura 85 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 6

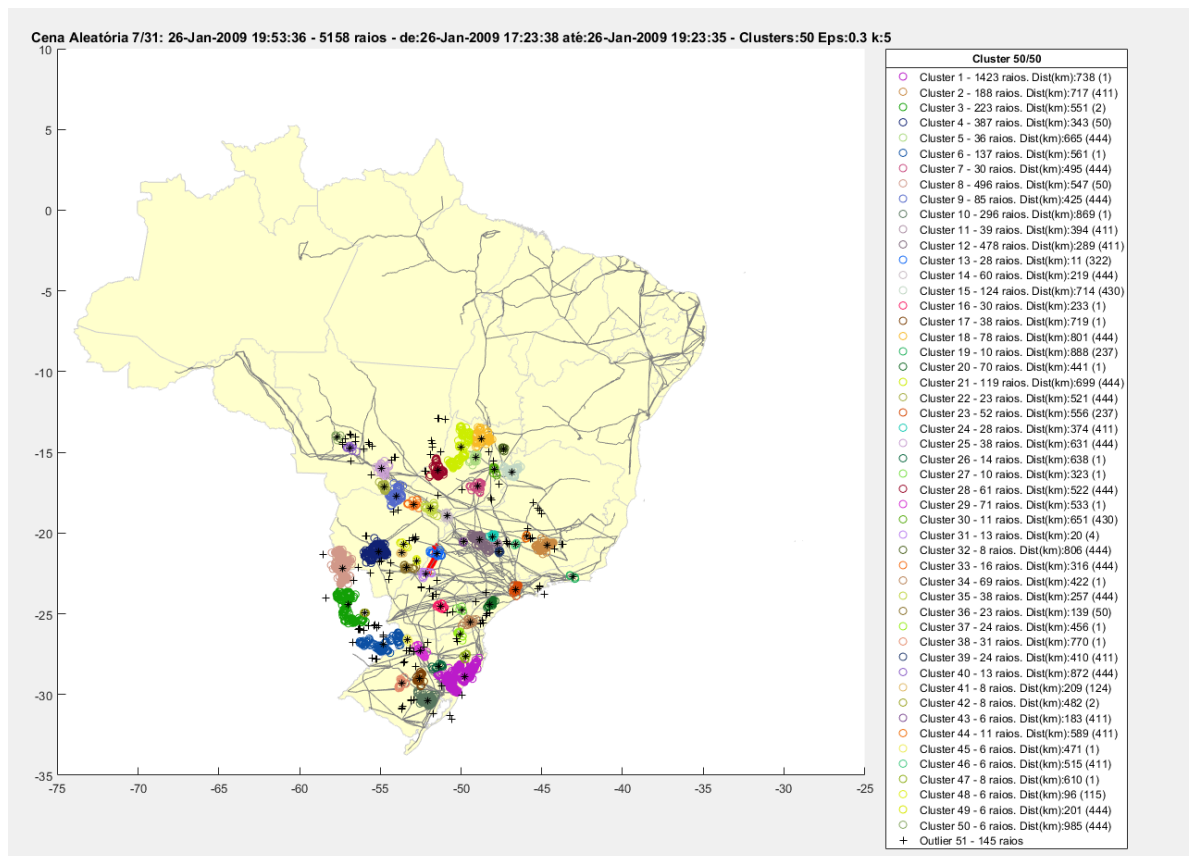


Figura 86 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 7

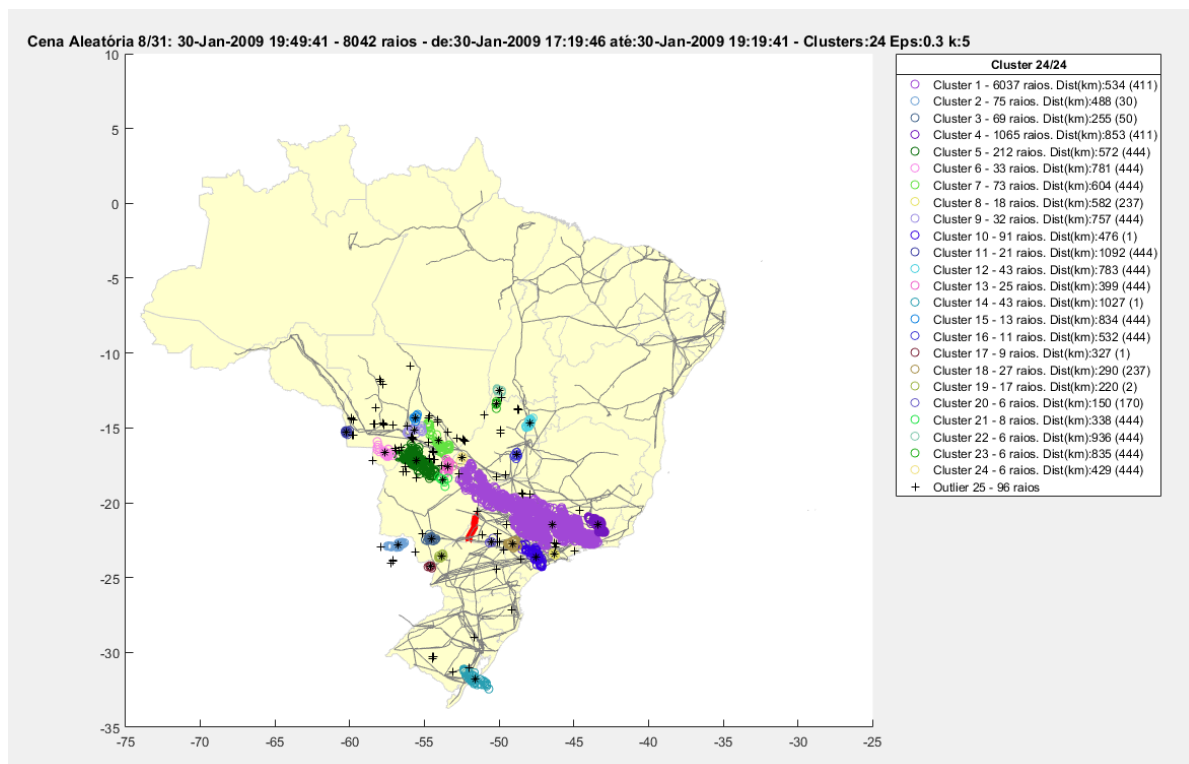


Figura 87 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 8

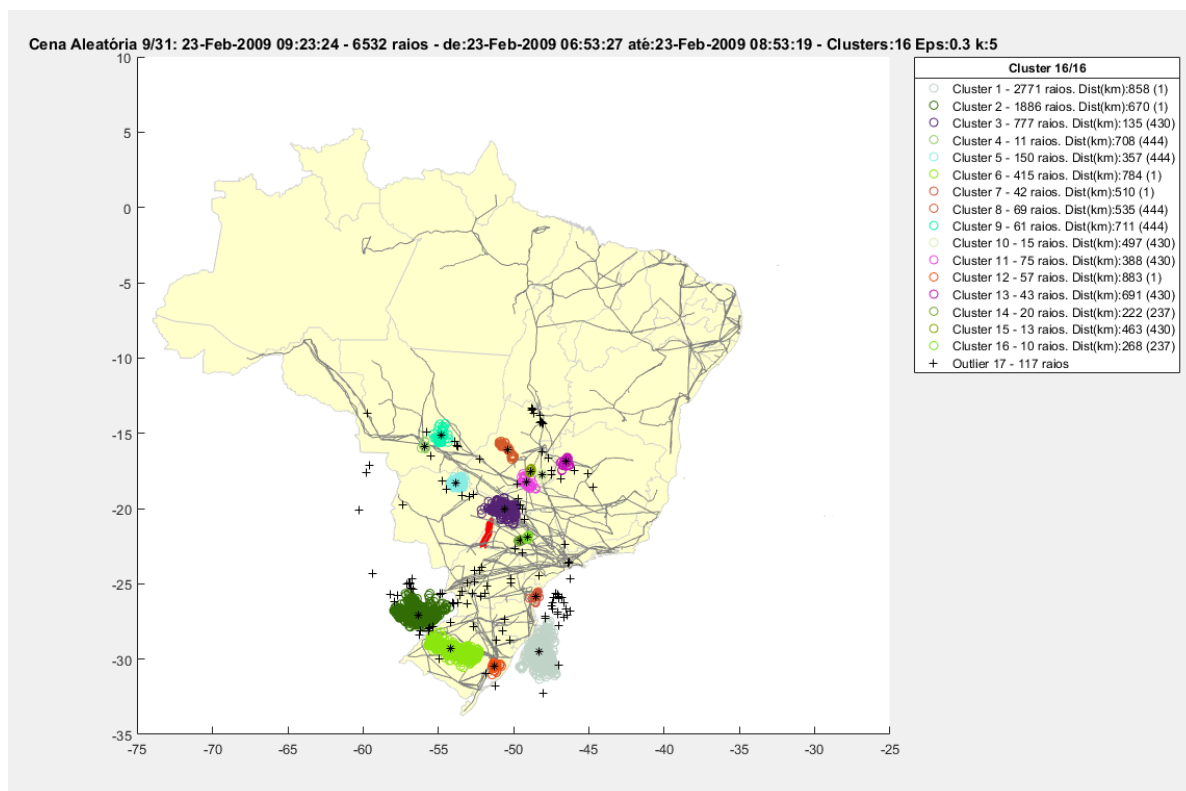


Figura 88 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 9

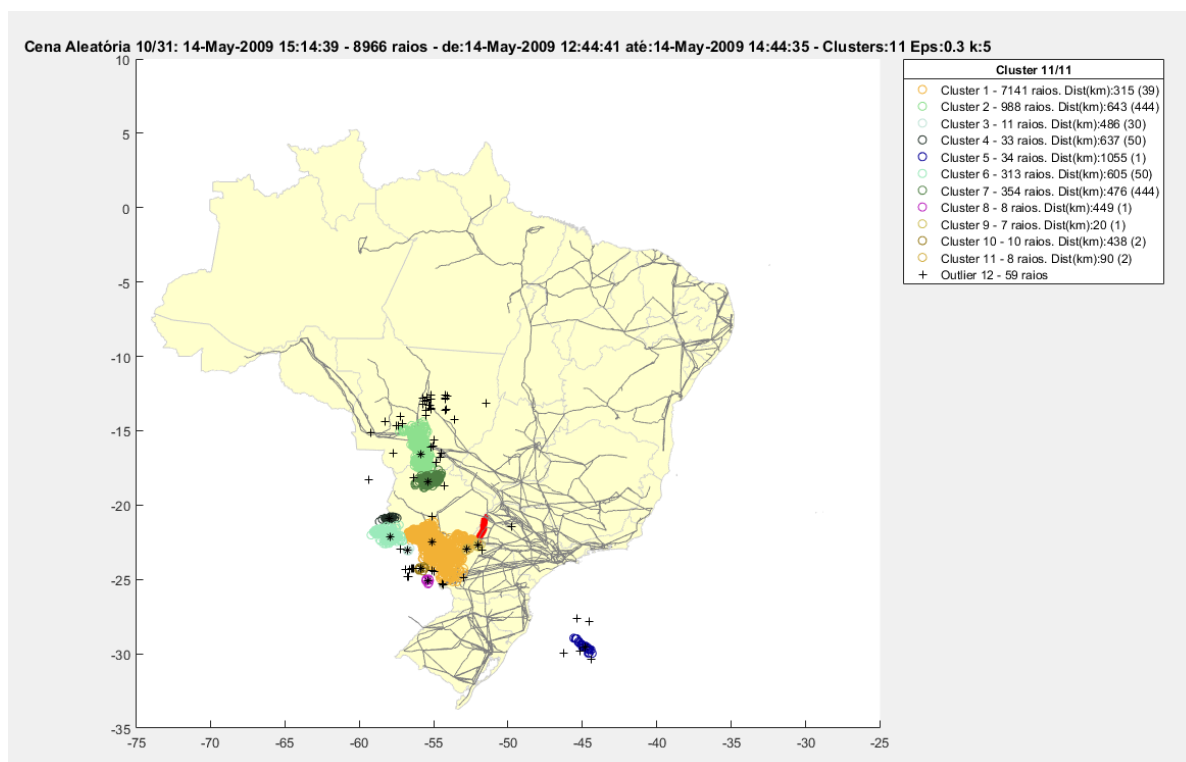


Figura 89 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 10

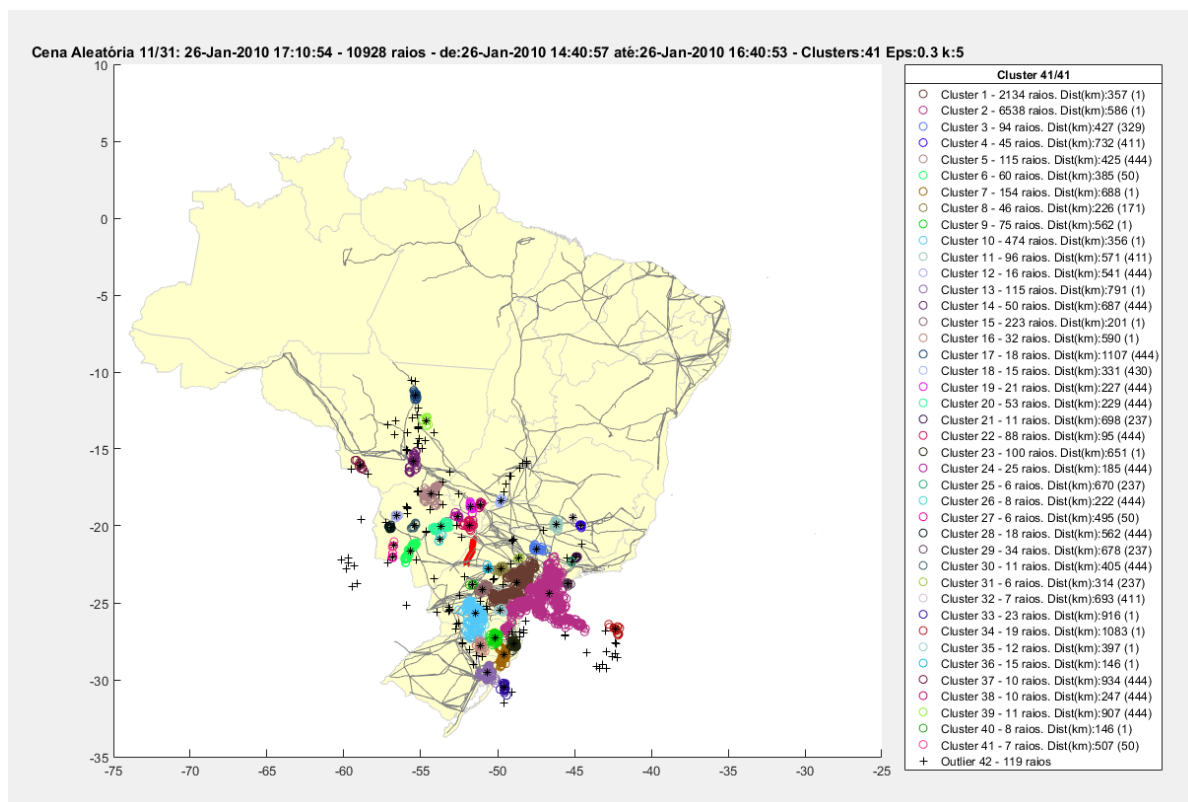


Figura 90 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 11

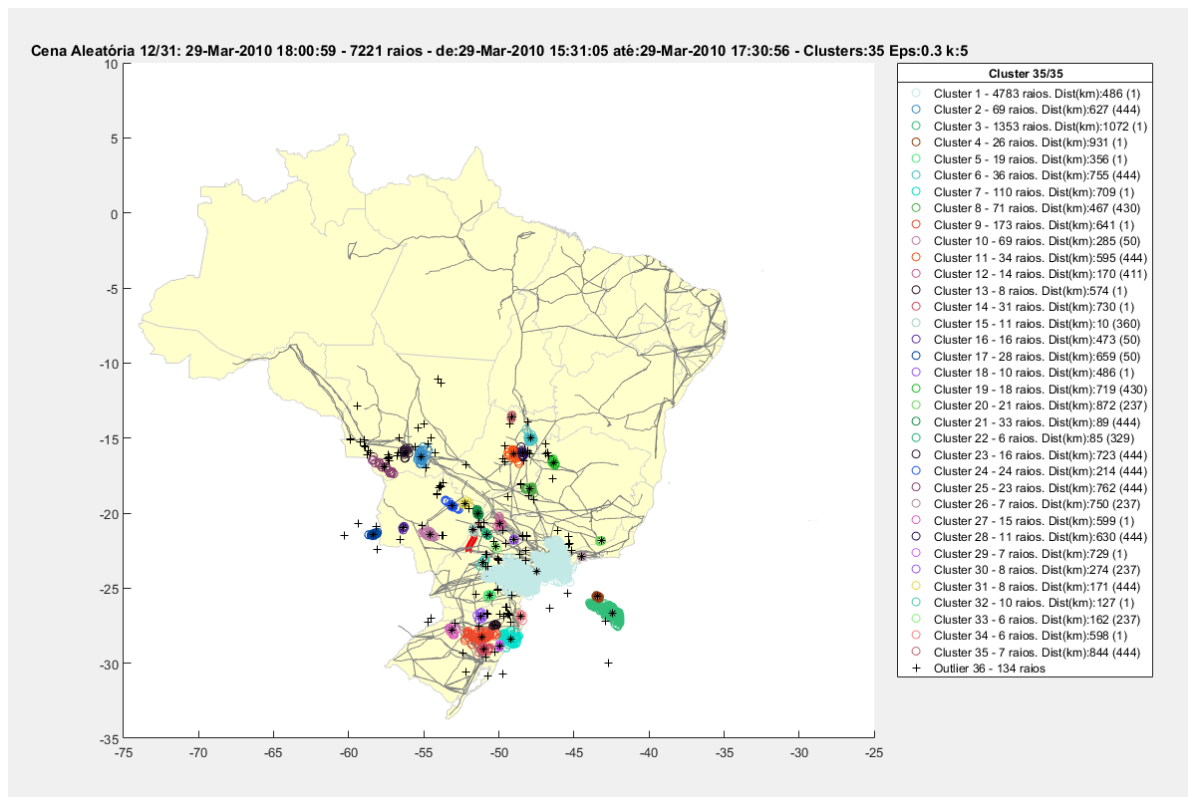


Figura 91 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 12

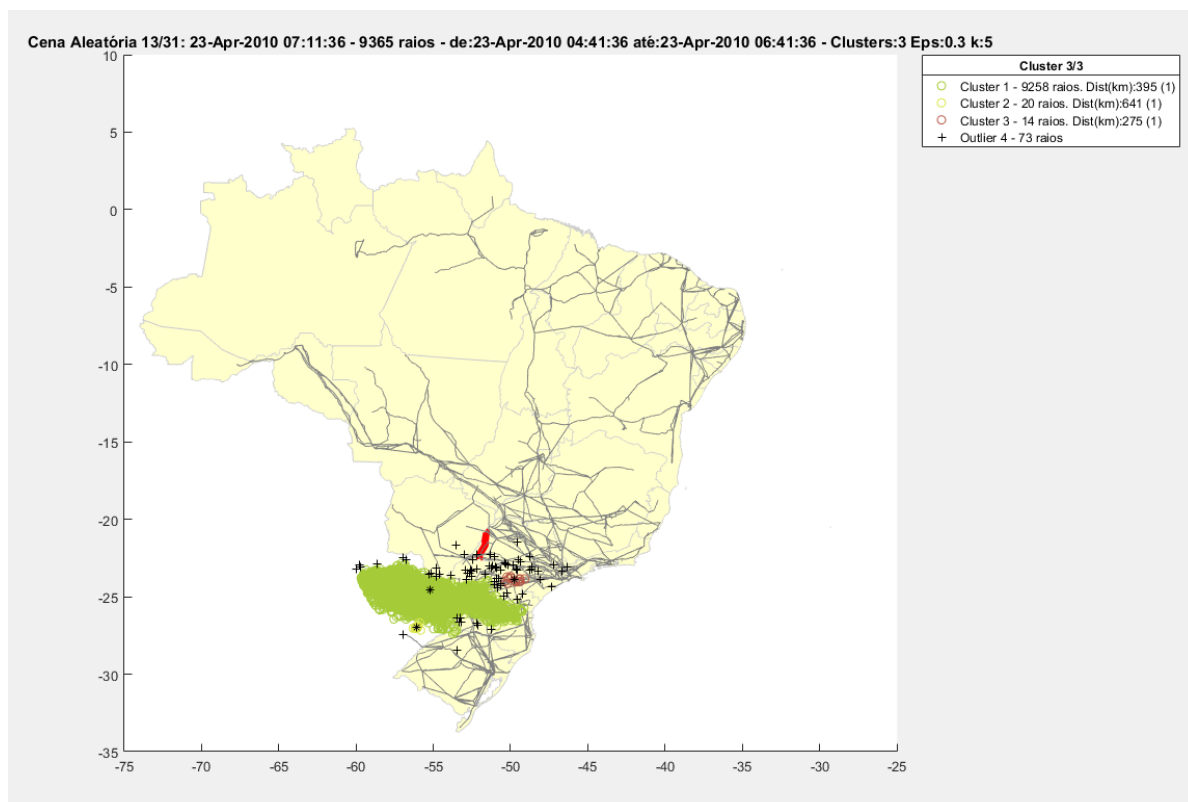


Figura 92 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 13

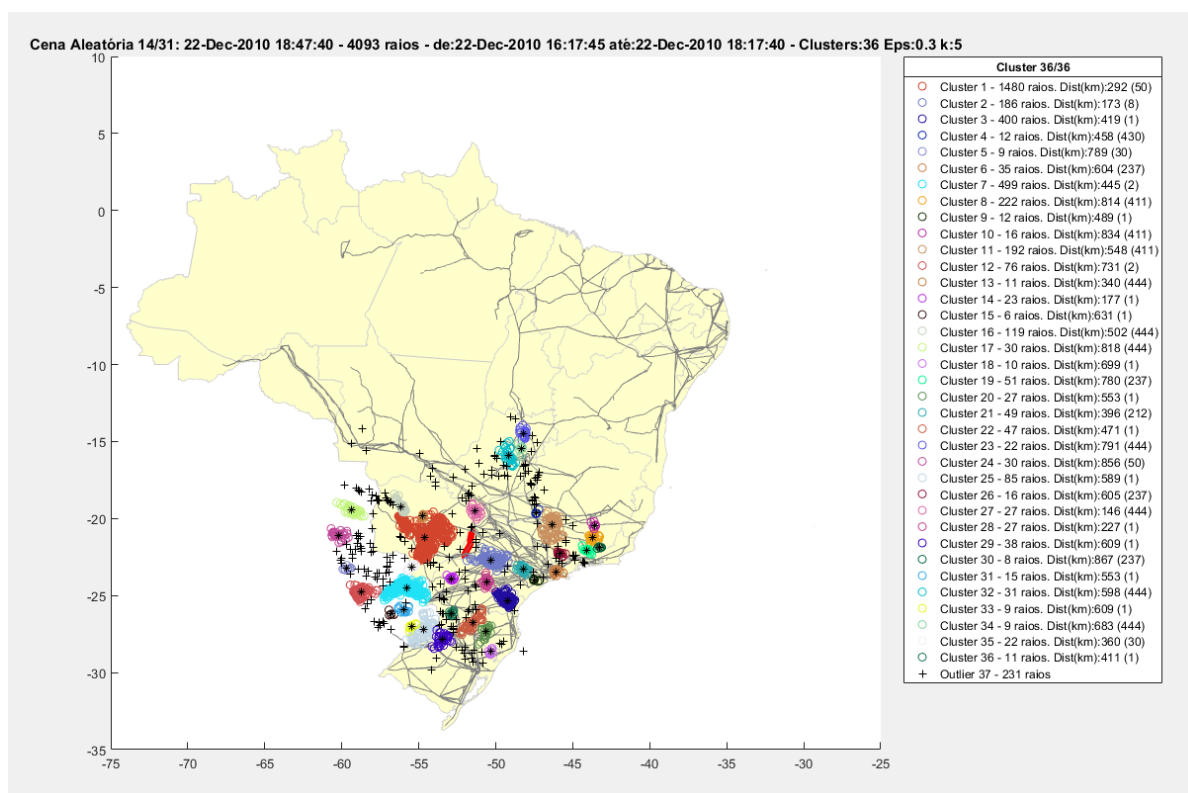


Figura 93 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 14

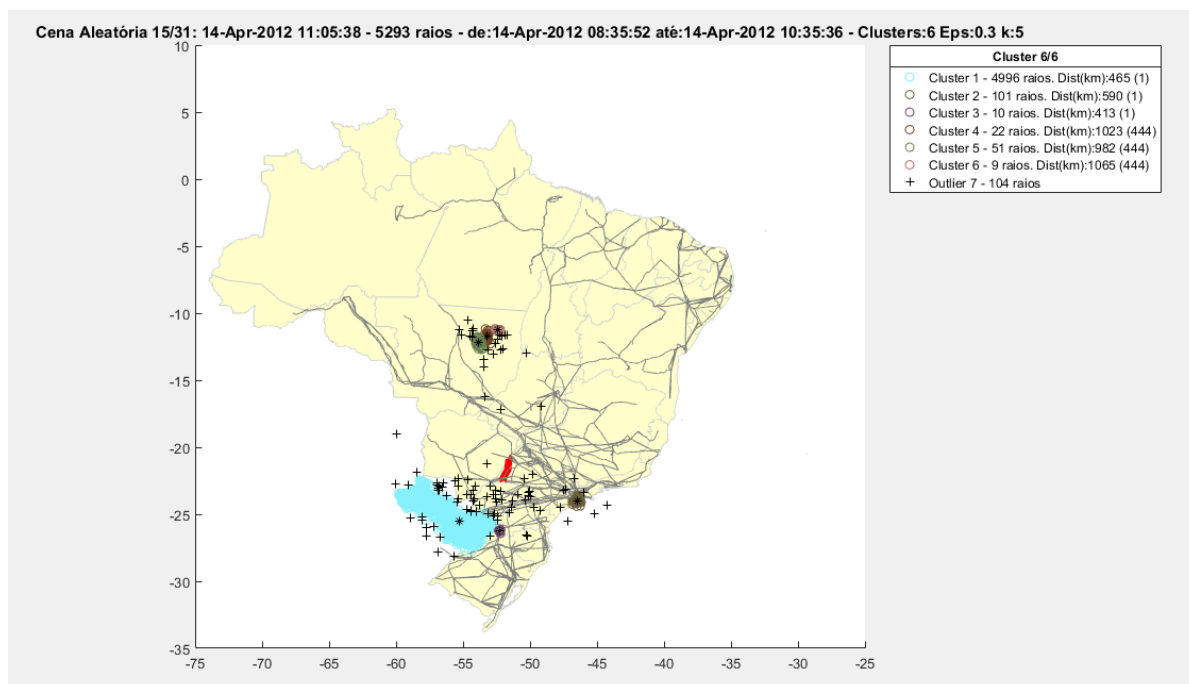


Figura 94 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 15

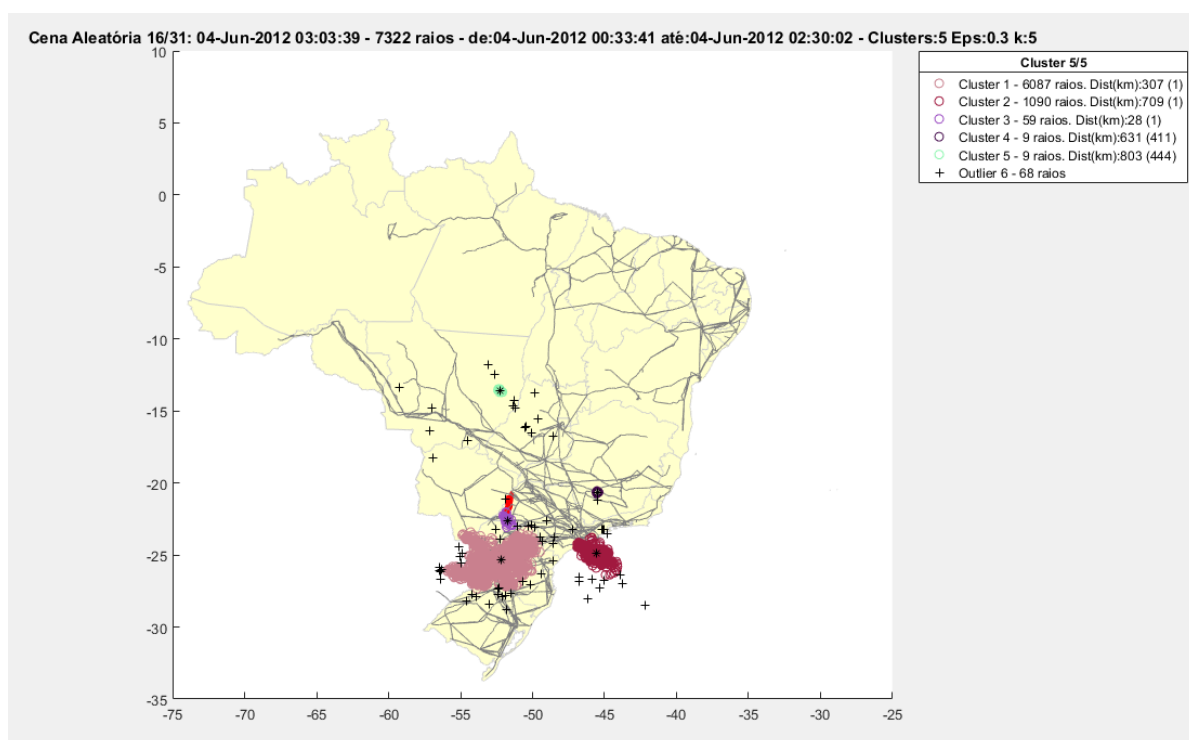


Figura 95 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 16

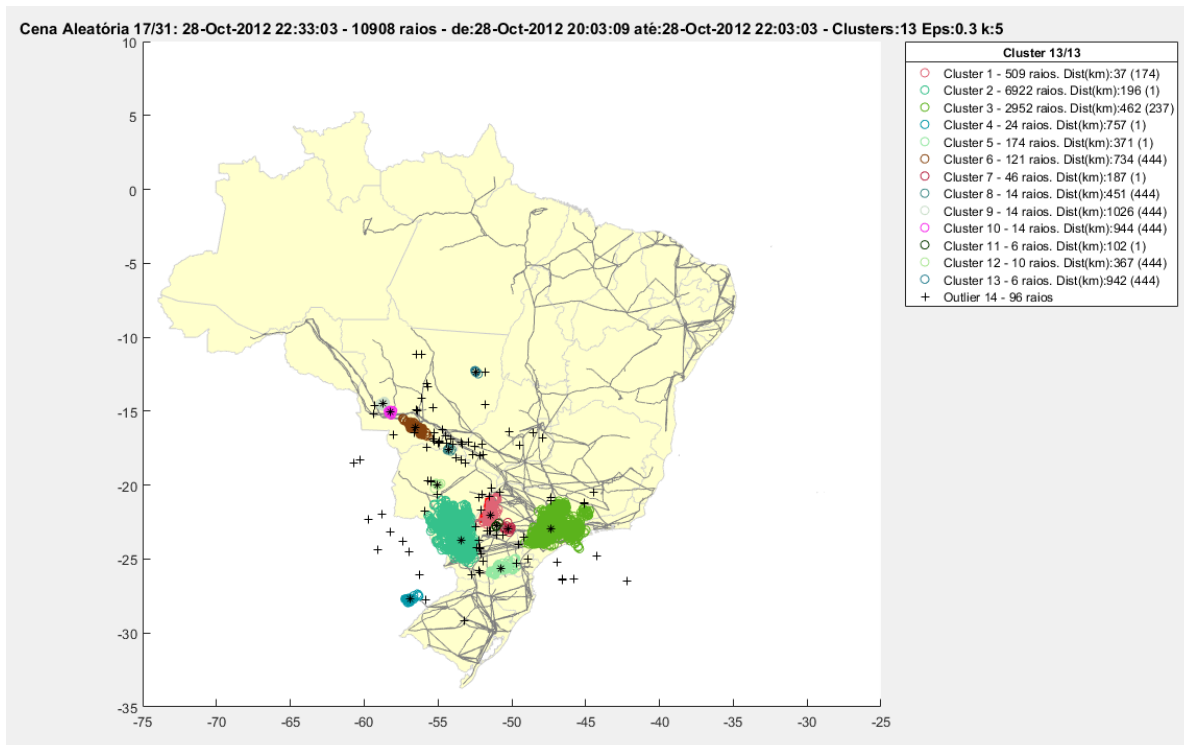


Figura 96 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 17

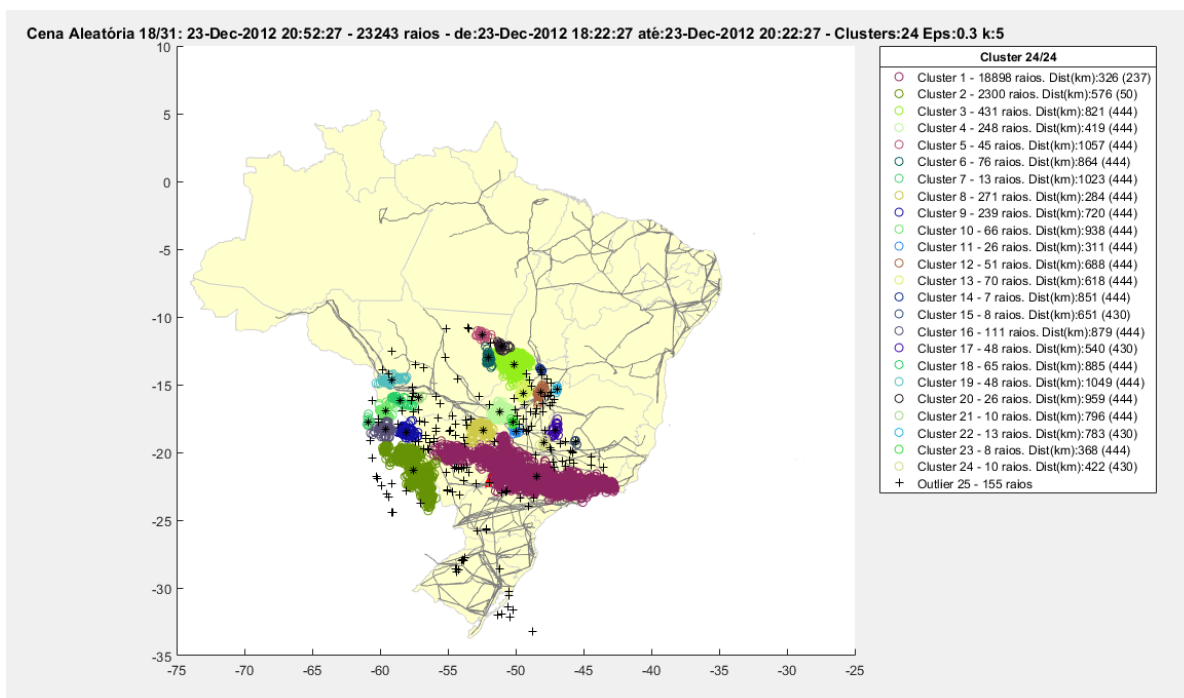


Figura 97 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 18

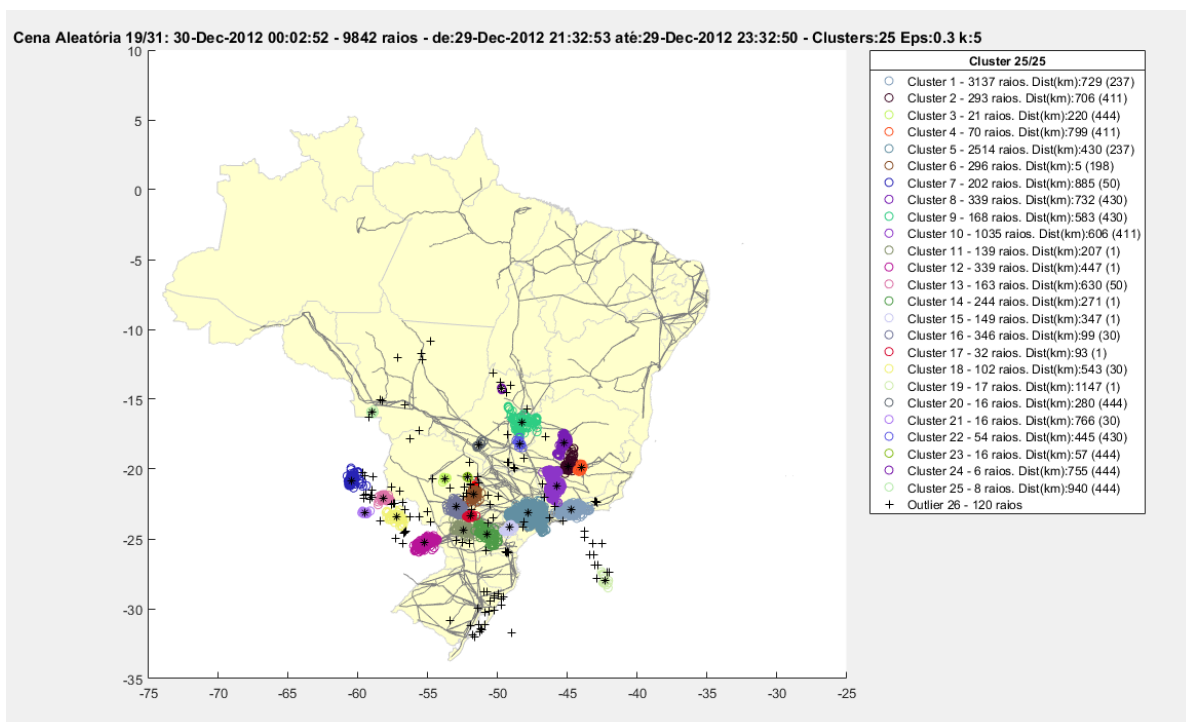


Figura 98 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 19

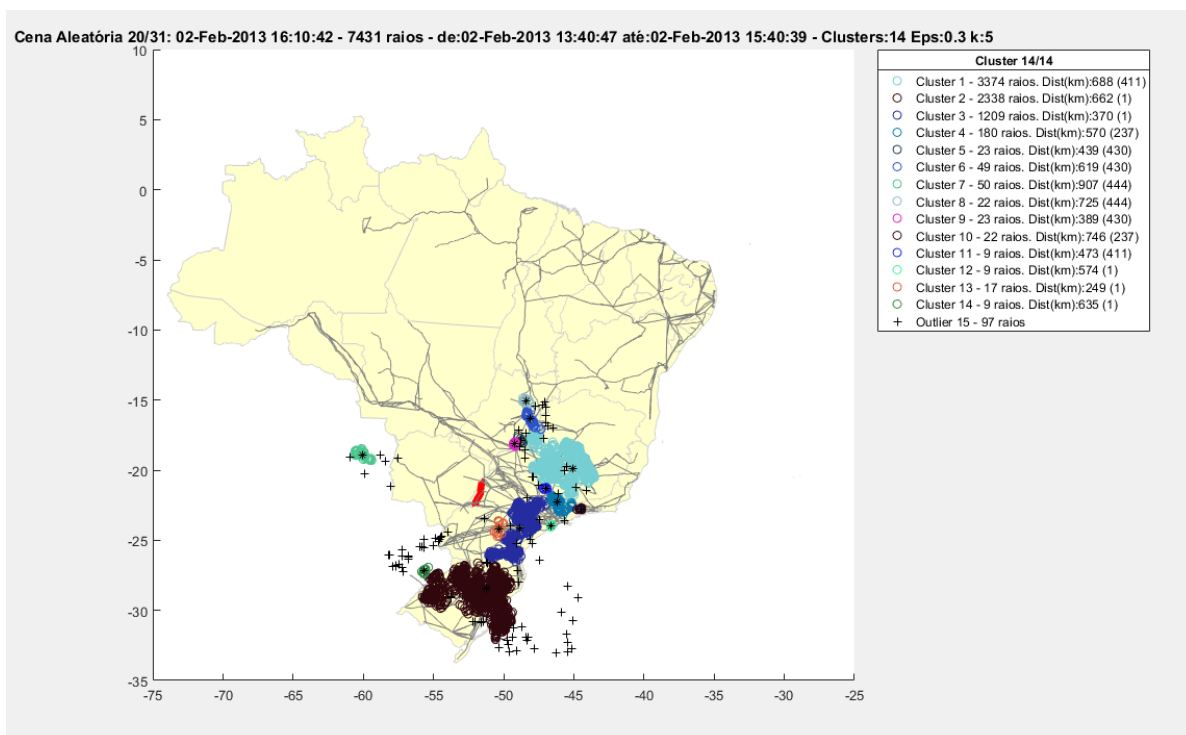


Figura 99 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 20

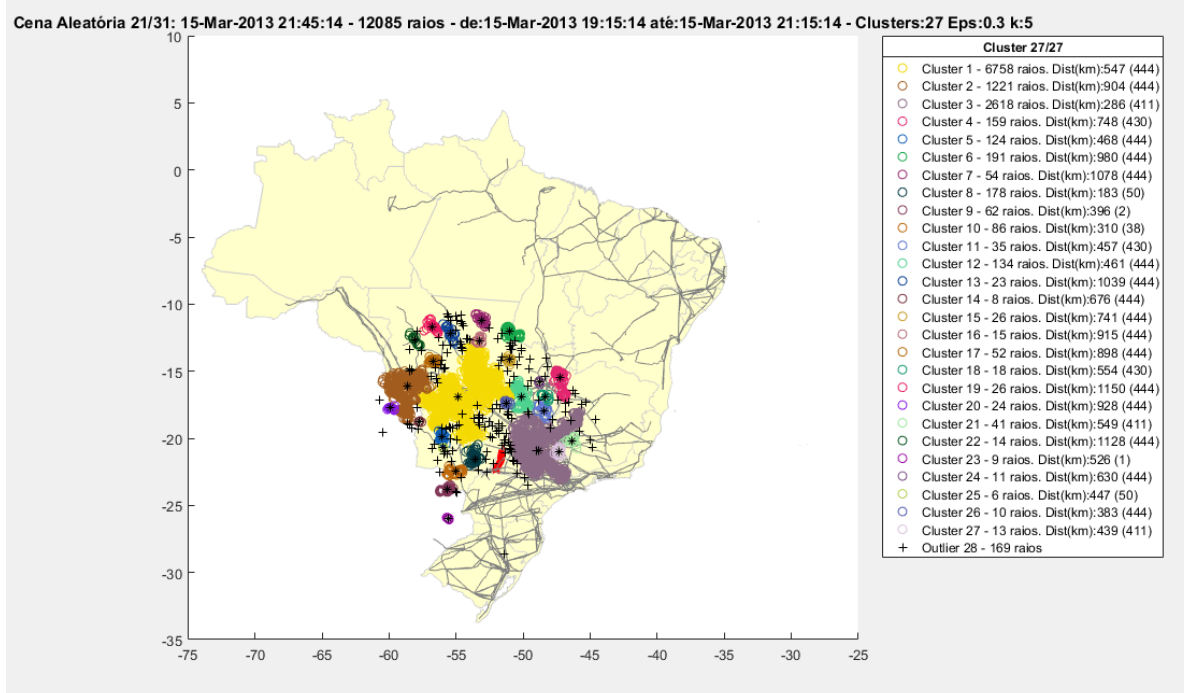


Figura 100 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 21

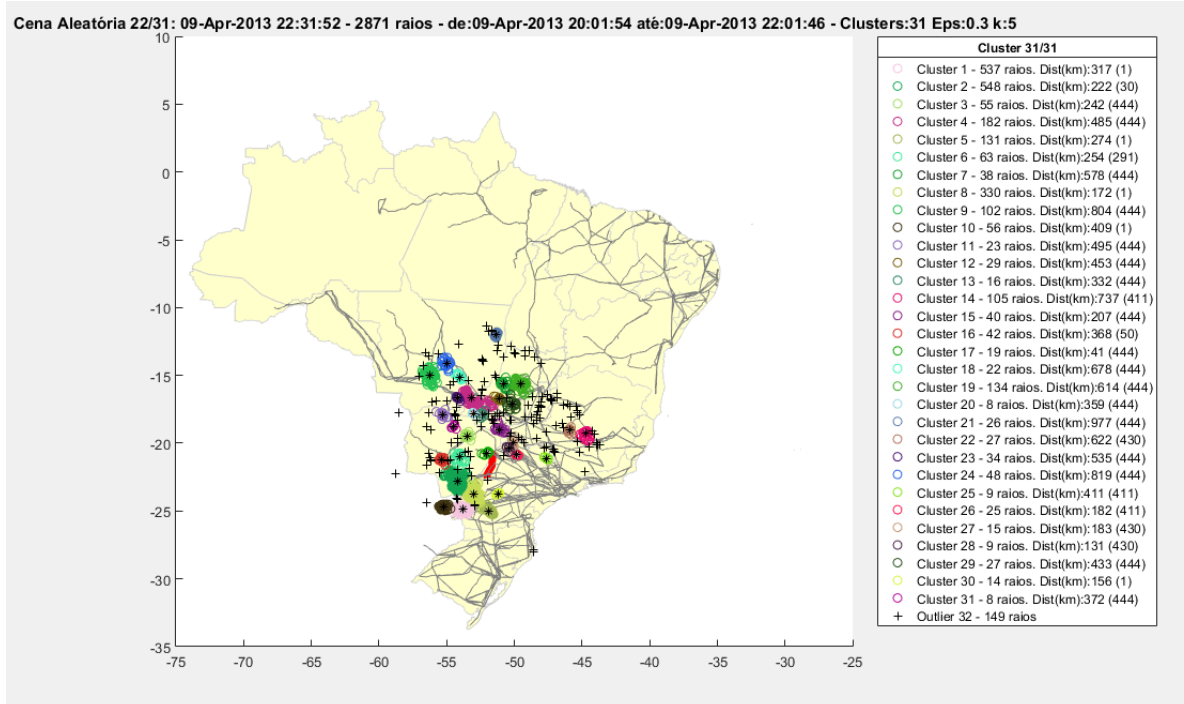


Figura 101 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 22

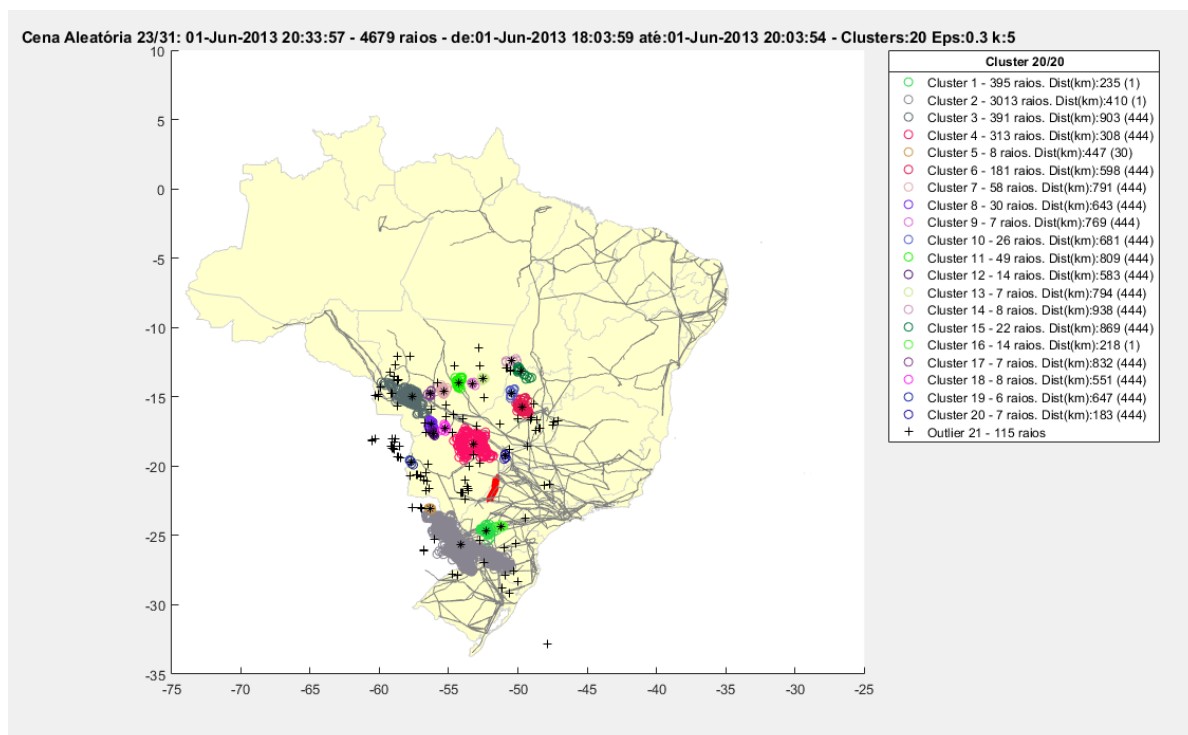


Figura 102 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 23

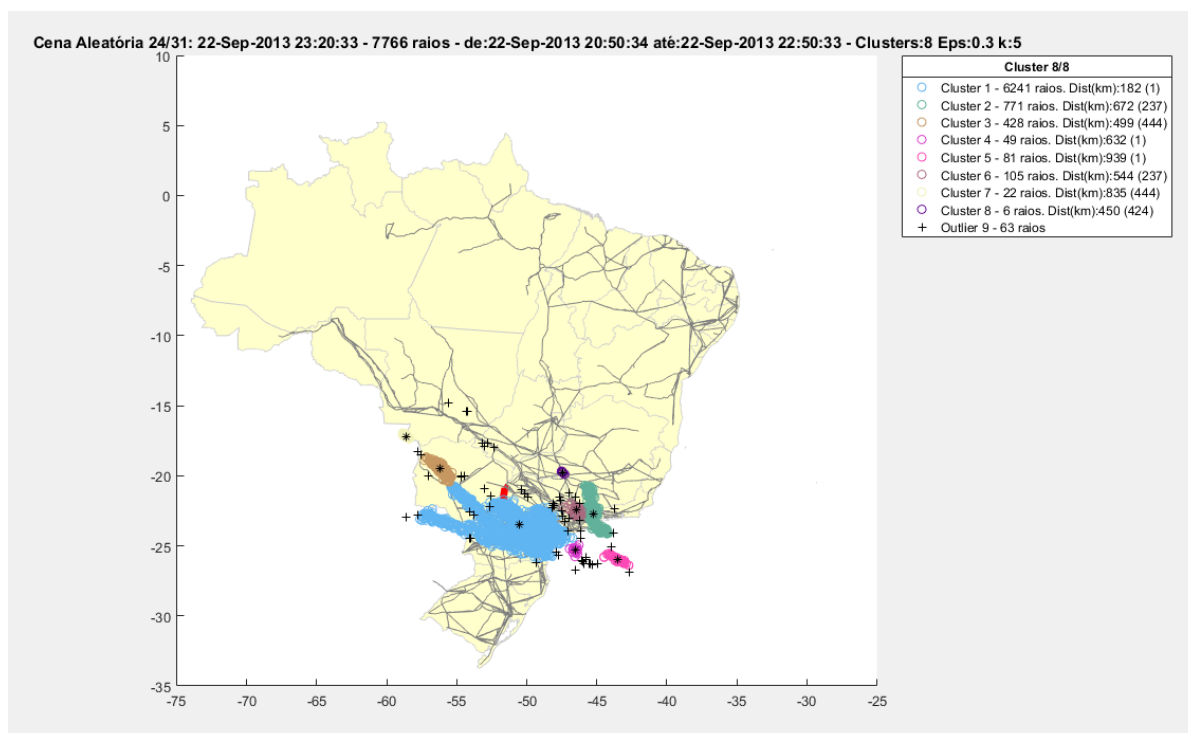


Figura 103 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 24

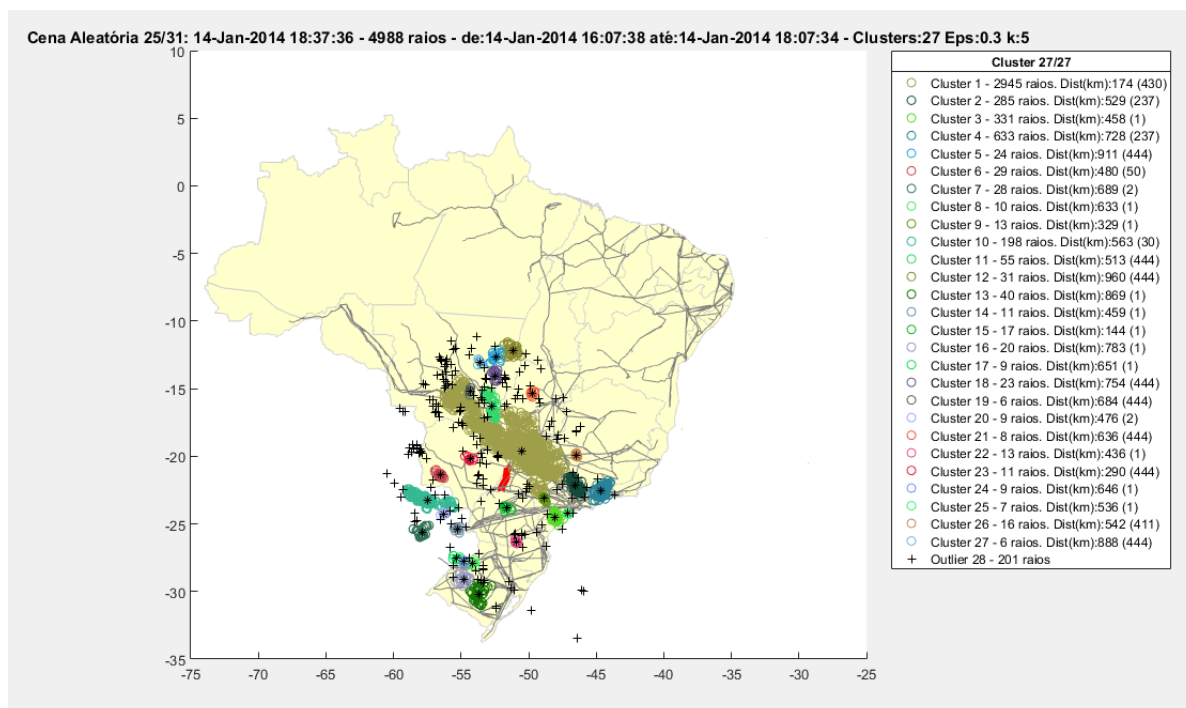


Figura 104 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 25

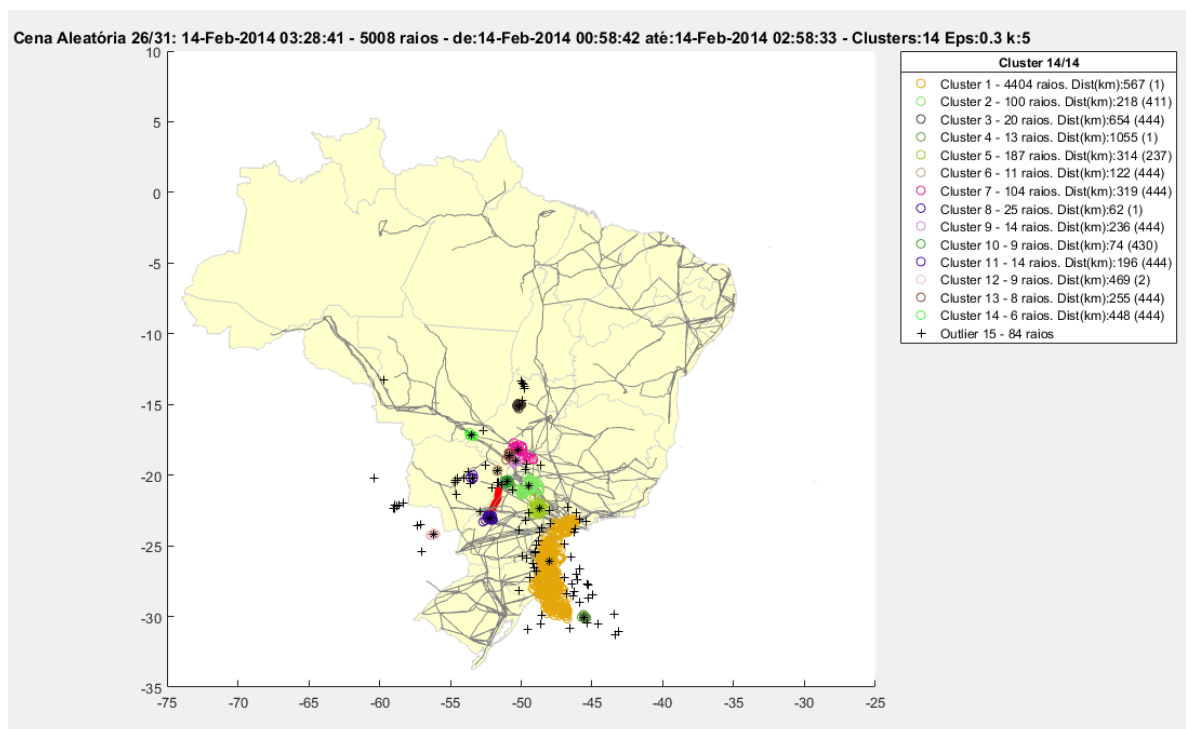


Figura 105 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 26

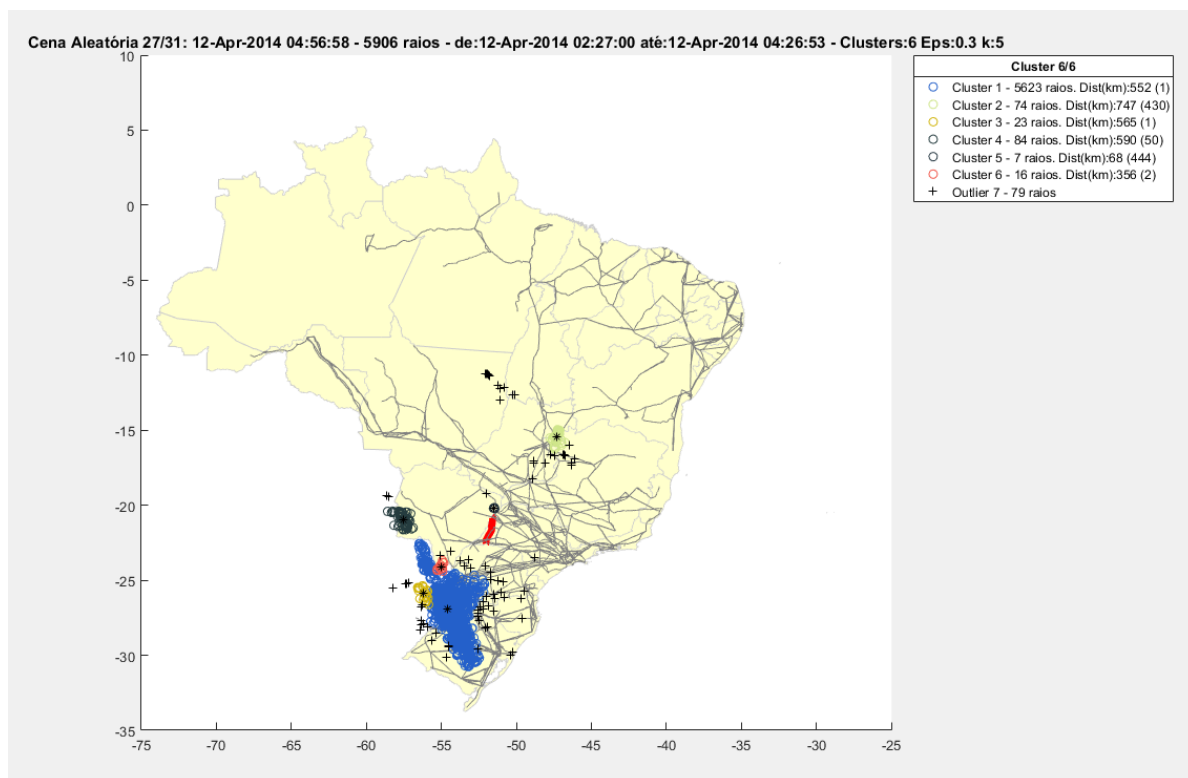


Figura 106 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 27

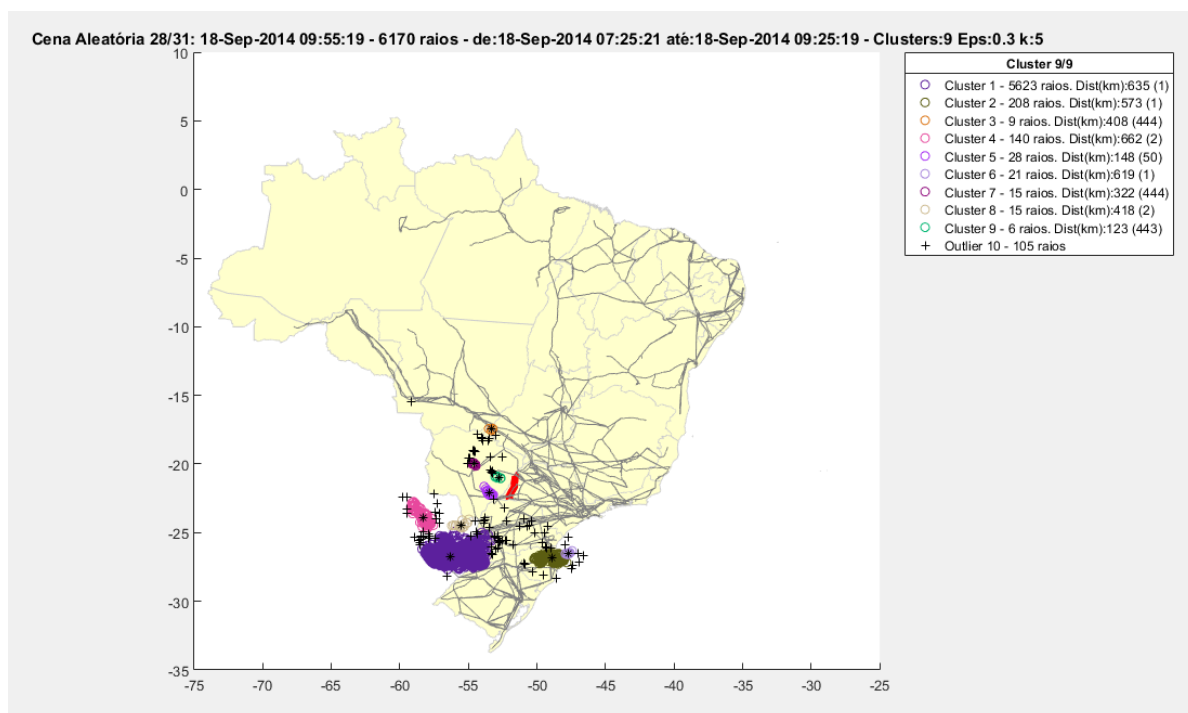


Figura 107 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 28

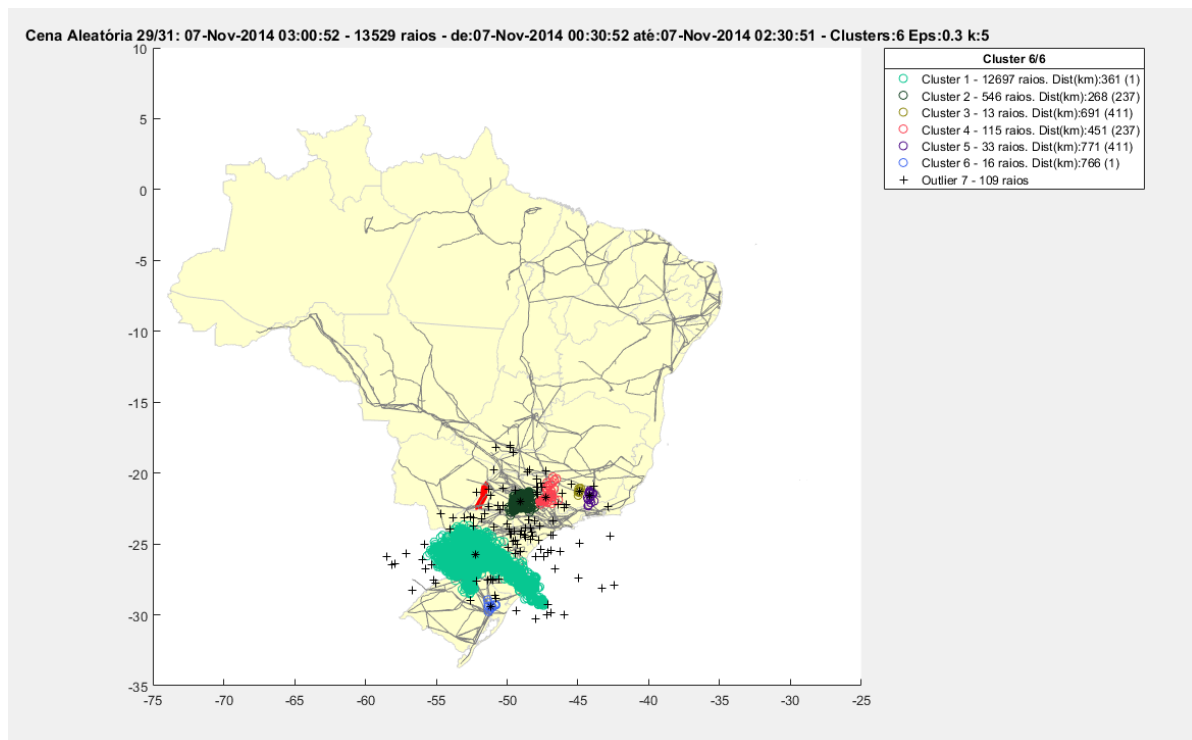


Figura 108 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 29

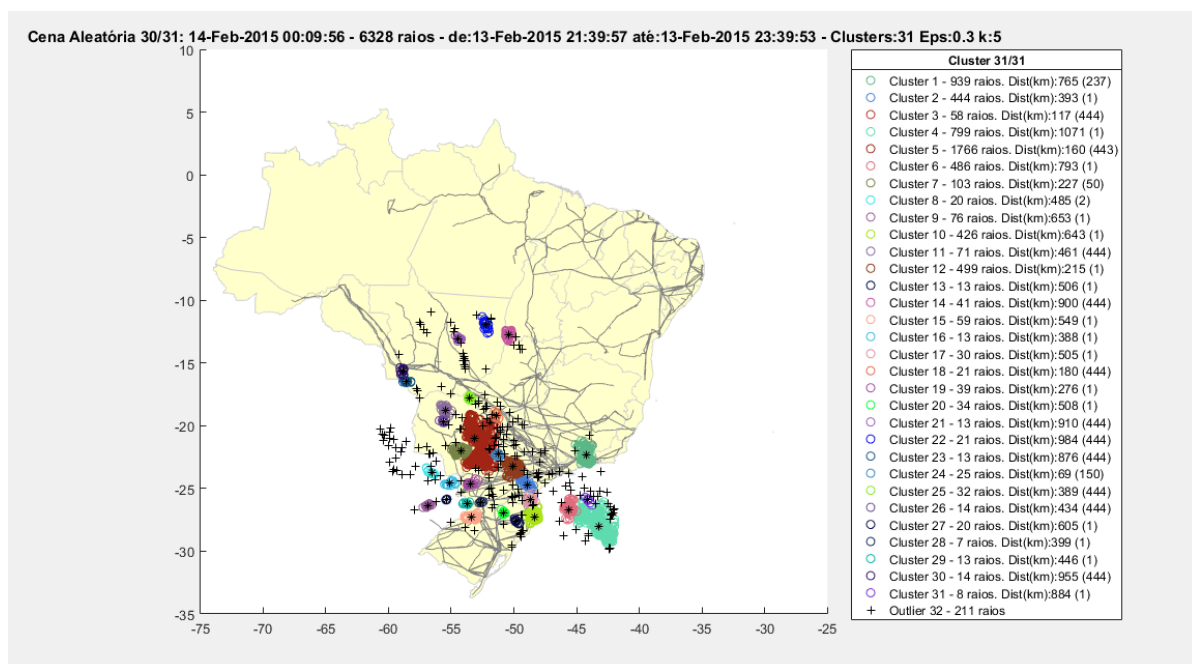


Figura 109 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 30

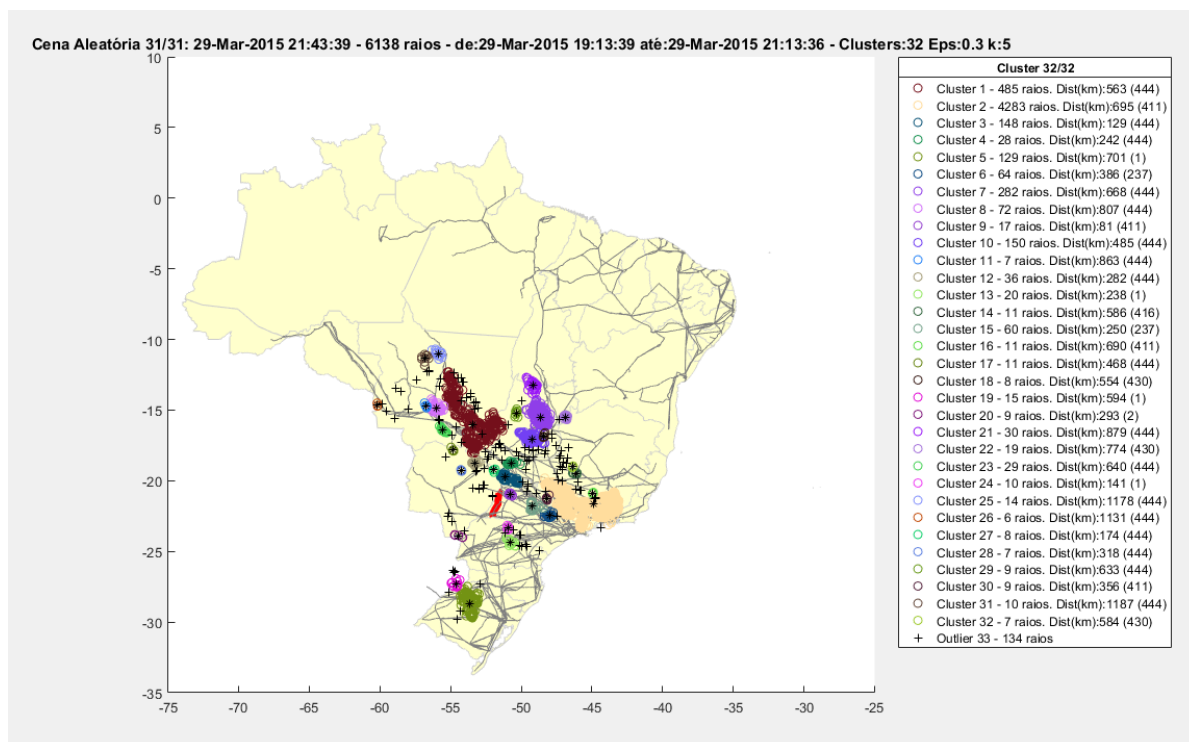


Figura 110 - Agrupamento das descargas atmosféricas não associadas ao desligamento da LT 440 kV TAQUARUÇU/JUPIA – cena 31